



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

# 챗봇과 온라인 설문 조사의 설문 문항에 따른 응답 품질 비교 분석:

－ 객관적 사실, 주관적 의견 및 개인정보  
질문을 중심으로 －

2020년 8월

서울대학교 대학원

융합과학부 디지털정보융합전공

배 범 근

# 챗봇과 온라인 설문 조사의 설문 문항에 따른 응답 품질 비교 분석:

- 객관적 사실, 주관적 의견 및 개인정보  
질문을 중심으로 -

지도 교수 서 봉 원

이 논문을 공학석사 학위논문으로 제출함  
2020년 7월

서울대학교 대학원  
융합과학부 디지털정보융합전공  
배 범 근

배범근의 공학석사 학위논문을 인준함  
2020년 7월

위 원 장 \_\_\_\_\_ 이 교 구 \_\_\_\_\_ (인)

부위원장 \_\_\_\_\_ 이 준 환 \_\_\_\_\_ (인)

위 원 \_\_\_\_\_ 서 봉 원 \_\_\_\_\_ (인)

## 초 록

설문조사는 모집단의 속성을 가장 잘 이해할 수 있게 해주는 중요한 방법들 중에 하나이다. 최근에는 온라인 설문조사가 가장 많이 사용되기는 하지만 여전히 낮은 응답률이나 응답품질이 떨어지는 문제점이 있다. 본 논문에서는 설문조사의 새로운 방법으로서 주관적 의견, 객관적 사실 및 개인정보와 같은 질문의 종류에 따라 챗봇이 유용성이 있는지에 대해 학문적으로 연구해보고 챗봇을 통한 설문과 기존의 온라인을 이용한 설문의 차이점에 대해 비교 실험을 하여 분석을 하였다.

이러한 연구목표를 달성하기 위해 Iterative design process를 이용하여 설문조사를 위한 챗봇 prototype을 만들어 여러 번의 실험을 통해서 받은 피드백을 통해 디자인을 개선하도록 하였다. 챗봇은 fluently.ai 플랫폼과 Chatfuel 플랫폼을 사용하여 제작하였으며 페이스북 메신저를 기반으로 운용하도록 설계되었다.

여러 번에 걸친 User Study를 통해 부드러운 말투나 감정의 교감 등의 Human Nature와 관련된 요소들이 설문조사의 결과에 긍정적인 기여를 하는 것으로 파악이 되었다. 이러한 특징을 챗봇의 주요한 요소로 정의하고 온라인 설문조사와의 비교 실험을 통해 질문 종류에 따라 응답에 대한 No Response, 응답 길이 및 Differentiation에 대해 정량적인 결과를 바탕으로 챗봇 설문조사가 온라인 설문조사보다 객관적 사실 및 주관적 의견에서 응답 품질의 차이가 있음을 확인할 수 있었다. 사후 인터뷰를 통해 챗봇 설문조사에 대한 정성적인 분석도 진행이 되었으며 그 결과로 챗봇을 통해 설문조사에 대한 몰입감이 높아졌으며 그 유용성도

확인 되었다.

본 연구를 통해 챗봇이 설문조사의 한 가지 중요한 방법론으로 사용될 수 있으며 그 결과가 효과적일 수 있음을 보여주는 이론적인 토대를 마련할 수 있게 되었고 주관적인 의견이 필요한 설문에 대해서는 챗봇이 온라인 설문조사에 비해 더욱 효과적임을 알수 있었다. 그리고 여러 번의 실험을 통해 파악된 챗봇을 설문조사에 응용하기 위해 참고할 만한 몇 가지 디자인 고려사항을 제시한 것에도 의의가 있다.

주요어: 챗봇, 설문조사, 응답품질, 주관적 의견, 객관적 사실

학번: 2017-26601

# 목 차

제 1 장 서론 .....	1
제 1 절 연구의 배경 .....	1
제 2 절 연구의 동기 .....	4
제 3 절 연구의 목표 .....	7
제 4 절 연구의 방법 .....	9
제 5 절 연구의 기여점 .....	11
제 2 장 선행연구 .....	12
제 1 절 챗봇(Chatbot) .....	12
제 2 절 다양한 설문조사 방법에 대한 연구 .....	15
제 3 절 설문조사의 평가 방법 .....	20
제 3 장 연구 문제 .....	22
제 1 절 연구 문제 .....	22
제 4 장 연구 방법 .....	25
제 1 절 연구방법 요약 .....	25
제 2 절 User Study #1 (챗봇 설문의 인식 조사) .....	28
제 3 절 User Study #2 (챗봇 설문조사 실험) .....	32
제 4 절 Evaluation .....	38
제 4 장 연구 결과 .....	47
제 1 절 정량적 분석 .....	47
제 2 절 정성적 분석 .....	58
제 3 절 챗봇 설문 조사 디자인을 위한 고려 사항 .....	69

제 5 장 결 론 .....	78
제 1 절 연구 요약 .....	78
제 2 절 연구의 시사점 및 의의.....	81
제 3 절 연구의 한계 및 Future Work.....	84
참 고 문 헌.....	86
Abstract.....	95

## 그림 목차

그림 1. 2가지 다른 설문조사 방법에 대한 예시 .....	27
그림 2. 챗봇을 통한 설문조사의 예시 동영상 .....	28
그림 3. 챗봇이 온라인 설문조사보다 개선될 것인가에 대한 설문 결과 .	29
그림 4. 카카오뱅크 설문에 대한 챗봇 화면 .....	33
그림 5. 챗봇 설문 사용 후 응답률 및 응답품질 예측의 변화 .....	35
그림 6. 온라인 설문조사 항목 .....	39
그림 7. 챗봇 설문조사 항목 .....	40
그림 8. 설문조사 타입 및 주제별 실험 assignment 방법 .....	42
그림 9. 챗봇과 웹의 주관적 의견 / 주관식의 Length .....	52
그림 10. 챗봇과 웹의 주관적 의견 / 주관식의 No Response .....	52
그림 11. 챗봇과 웹의 객관적 사실 / 객관식의 Differentiation .....	53
그림 12. 챗봇과 웹의 객관적 사실 / 객관식의 No Response .....	53
그림 13. 챗봇과 웹의 주관적 의견 / 객관식의 No Response .....	54
그림 14. 챗봇과 웹의 객관적 사실 / 주관식의 No Response .....	54
그림 15. 챗봇 설문조사에 대한 거부감 .....	56
그림 16. 설문 시 모바일 키보드 사용의 불편함 조사 .....	67
그림 17. 설문시 화면 사이즈 제약으로 인한 스크롤의 불편함 .....	67
그림 18. NH수준의 메시지 조작 예시 [5] .....	72
그림 19. 의인화 조작 예시 [4] .....	73



## 표 목차

표 1. 각 설문조사 방식의 장/단점 [82] .....	15
표 2. 결과의 정규성 검증 .....	48
표 3. 챗봇과 웹의 응답 품질 차이 검증 결과 .....	51
표 4. 챗봇과 웹의 유용성과 몰입감 비교 .....	55

# 제 1 장 서 론

## 제 1 절 연구의 배경

설문조사는 모집단의 속성에 대해 이해하기 위해 개인으로부터 정보를 수집해내는 체계적인 방법[18]으로 학문적인 연구를 진행함에 있어 데이터를 얻거나 실험을 통한 결과를 만들어 내기 위해서는 없어서는 안 되는 아주 중요한 방법 중에 하나이다. 초기의 설문조사는 인구통계를 위해서 사용되었지만 차츰 사회적 이슈를 이해하기 위해서 설문이 진행되기 시작하였고 이후 여론이나 시장조사를 위해 우편이나 전화와 같은 매체를 이용하는 등의 설문조사로 진화하였다. [13] 인터넷 기술의 발달과 보급의 확산으로 온라인으로 설문조사를 하기 시작하였고[17] 이후 휴대폰이 대중화 되면서 이를 위한 설문조사의 다양한 연구[9, 20, 21, 22]나 기술들이 나오게 되었다. 산업계에서도 기업의 정책이나 마케팅을 위한 수단으로서 설문 조사가 많이 이용되고 있으며 Global Market Research (ESOMAR)에 따르면 2014년에는 설문조사의 26%정도가 온라인을 이용하고 있고 특히 온라인 설문을 위한 소프트웨어의 시장은 계속 성장추세에 있으며 서베이몽키 등의

많은 기업들이 운영되고 있다.<sup>1</sup>

온라인 설문조사의 경우 기존의 설문조사 방식에 비해 비용이 적게 들고 편리하며 조사기간이 오래 걸리지 않은 장점이 있고[24, 45] 사용자로서도 원하는 시간에 설문을 진행 할 수 있으며 물리적인 인터뷰어가 없기 때문에 이로 인해 발생할 수 있는 오차도 줄어드는 등의 장점을 가지고 있다. [17] 한발 더 나아가 Mixed Survey Method[37, 38]나 Adaptive Survey[39, 40, 41] 등 기존의 설문조사 방식을 변경하고 조합하여 사용되고 있기도 한다. 하지만 온라인 설문조사의 경우 응답률이 기존 방식에 비해 떨어진다는 단점은 여전히 존재하고 있어[32, 35, 36, 17, 46] 응답률이나 응답품질을 높이는 디자인을 위한 다양한 연구[26, 27, 28, 29, 30, 31, 33, 34] 등이 꾸준히 진행되어 오고 있지만 새로운 방법을 찾기 보다는 인센티브를 주거나 여러 번 알림을 주는 등의 온라인 설문조사 내부의 정책을 바꾸는 것들에 의존하고 있다.

최근에는 여러 분야에서 널리 사용되고 있는 챗봇을 설문 조사에 접목하여 실험하는 연구들이 많이 진행되고 있다. 인터넷이 널리 보급 되면서 새롭게 등장한 것이 온라인 설문 조사이었던 것처럼 발전된 AI 기술을 접목한 챗봇 설문 조사가 서베이의 패러다임으로 변경시키고 새로운 설문 조사의 방법론으로 자리를 잡을 수 있을지에

---

<sup>1</sup> Online Survey Software in the US Market Size 2002-2025, IBISWorld

<https://www.ibisworld.com/industry-statistics/market-size/online-survey-software-united-states/>

대한 중요한 시점이라 보여진다. 전통적인 설문 조사들이 가지고 있는 단점들 중에 특히 온라인 설문조사에서 문제가 되고 있는 응답 품질에 대해서 챗봇 설문조사를 통해 개선될 수 있는지에 대해서 다양한 방법으로 연구가 진행되고 있다. 대화형 챗봇의 경우 메신저 프로그램을 통해 진행이 되기 때문에 흔히 말하는 친근한 말투[2]로 마치 친구와 얘기하듯이 설문을 진행할 수 있는 장점이 있고 이러한 접근이 사용자에게 거부감이 있는지, 그리고 주관적 의견에 대한 질문뿐만 아니라 객관적 사실, 나아가서 개인정보에 대한 질문에 대해서도 주관식, 객관식에 대해 각각 응답 품질이 좋아지는지를 본 연구에서 확인하고자 하였다.

## 제 2 절 연구의 동기

최근 인공지능기술의 발달로 인해 가장 활발히 연구되면서 매일 새로운 형태의 제품들이 쏟아지며 폭발적인 성장을 하고 있는 분야가 있다. 바로 대화형 에이전트인 챗봇이다. [14, 15, 16] 챗봇(Chatbot)은 사용자와 자연어로 소통하는 대화형 시스템으로 모바일 메신저나 PC상의 대화형 인터페이스를 이용하여 실제 상담원이나 제품 판매자들에 해당하는 사람이 대화[19]하는 것이 아니라 Rule Based Model, Retrieval Based Model 혹은 Generative Model 기반의 기술로 운영이 되는 프로그래밍된 대화형 에이전트를 의미한다. [42] 일반적으로 챗봇은 PC나 메신저상에서 사용할 수 있지만 점점 모바일 메신저가 일반화되는 추세에 있으며 모바일 메신저의 경우 전 세계적으로 월간 이용자수는 50억명이 넘으며 특히 페이스북은 왓츠앱(페이스북이 인수, 2014)을 포함하여 20억명 이상의 이용자를 보유하고 있고 스마트폰의 최고의 Killer App이라고 할 수 있다.<sup>2</sup> 2017년 F8 conference를 통해 새롭게 “Messenger Platform 2.0”을 발표했으며 월간 12억이 넘는 사용자와 10만개가 넘는 동작중인 챗봇을 위해 새로운 기능들을 선보였다.<sup>3</sup> 챗봇 시장의 경우 연평균

---

<sup>2</sup> 류한석. 2017. 챗봇 비즈니스 트렌드 및 시사점. 류한석 기술연구소 디지예코 보고서

<sup>3</sup> Mikhail Larionov. 2017. Messenger Platform 2.0 Debuts at F8. Facebook Developer News

20%이상의 성장을 예상하고 있으며 BFSI(Bank, Financial Services and Insurance)분야나 유통 및 e-커머스 분야에서 두드러지며 그 외에도 다양한 산업분야에도 적용되는 추세이다.<sup>4</sup> 또한 2014년에 열린 구글의 “Mobile First World Conference” 에서 처럼 이미 사람들이 컴퓨터나 태블릿에 비해서 모바일을 사용하는 시간이 훨씬 많으며<sup>5</sup> 온라인 서베이의 경우에도 모바일 디바이스를 사용하는 경우가 점점 증가하고 있고 이에 대한 연구도 이미 많이 진행되어 있고 [9, 10] 모든 산업과 기술은 모바일 디바이스에 적합한 환경을 꾸미는 것도 중요하게 여겨진다.

이러한 챗봇의 생태계가 가지고 있는 장점을 바탕으로 설문조사에 접목하려는 움직임도 많이 등장하고 있고 특히 페이스북 메신저를 기반의 설문을 위한 챗봇을 만들어주는 플랫폼인 ‘Surveybot’ 이 상용화 되어 있으며<sup>6</sup> 수많은 다른 챗봇 플랫폼들을 통해 챗봇 제작을 누구나 쉽게 할 수 있는 환경이 만들어지고 있다. 이러한 챗봇을 통한 설문조사에 대한 관심은 지속적으로 이루어져 왔고 최근에는 여러 연구 결과들이 발표가 되었다. [11, 2, 78, 80] 기존 실험들은 설문지 특징에 기반하기 보다는 일반적인 설문에서 그 응답품질을 확인하였기 때문에 어떤 설문인가에 따라 그 효과가 나타날 수 있을지를 일반화 하지는 못하고 있다. 본 연구에서는 기존 실험들의

---

<sup>4</sup> Grand View Research, Inc. 2017. Chatbot Market Size to Reach \$1.25 Billion by 2025 | CAGR: 24.3%. Grand View Research, Inc. 2017.

<sup>5</sup> 닐슨 코리아클릭. 2017. Mobile First를 넘어 Mobile Only로: Mobile Only 이용자 행태 분석. 뉴스레터 2017. 9. 26. 제 272-2호

<sup>6</sup> Surveybot. <https://surveybot.io/>

결과를 바탕으로 친근한 말투의 상호작용이 있는 챗봇 설문 조사를 진행했을 때 사용자의 거부감이 있는지를 확인하고 설문 문항의 특징에 따라 설문조사에 대한 응답 품질이 어떻게 차이가 나는지를 추가로 확인해보고 이에 대한 논의를 해보고자 한다. 그리고 챗봇 설문 조사를 진행할 때 필요한 가이드라인을 제시해주게 된다면 이후 연구자들에게 도움이 될 것이라 본다.

### 제 3 절 연구의 목표

설문을 위해 주어지는 설문 문항 형태는 크게 보면 “주관적 의견”과 “객관적 사실”에 대한 범주로 나뉜다. 국립국어원 표준국어대사전에 정의에 의하면 ‘의견’이라 함은 어떤 대상에 대하여 가지는 생각을 의미한다. 이는 자신의 생각을 표현하는 것이고 그것이 참인지 거짓인지는 누구도 판단할 수는 없다. 마찬가지로 ‘사실’은 “실제로 있었던 일이나 현재에 있는 일”을 말하거나 “객관적 현실에 부합하고 증거에 의해 참으로 증명될 수 있는 어떤 것”을 의미한다. 이와 같이 설문 조사의 목적인 데이터를 수집을 위해서는 위의 두 가지 항목에 해당하는 질문 내용을 던지도록 되어 있고 이에 대한 답변을 받고자 하는 것이다.

본 연구에서는 주로 모바일 메신저의 형태로 활용이 되어 간편함과 익숙함의 특징을 가지고 있는 챗봇이 객관식 및 주관식을 포함한 설문조사를 진행하는 새로운 도구로서 유용한지를 살펴보고 이를 페이스북 메신저에 기반한 “Chat Survey” [12]를 제작하여 온라인 설문조사와의 비교 실험을 통한 분석을 통해 챗봇을 사용함으로써 객관적 사실 및 주관적 의견의 설문에 대한 응답품질이 어떻게 차이를 나타나는지를 살펴보고 특별히 설문조사의 기본적인 통계의 기준으로 파악하는 개인정보에 대해서는 따로 범주화 하여 비교해보았다. 몰입감[43]과 유용성[44] 그리고 설문에 대한 거부감에 대해 파악하고



더불어 빠른 응답을 통한 신뢰성이나 응답 변별력에 따른 유효성을 확인하고자 하였다. [1, 67] 추가적으로 챗봇에 대한 디자인 제안 들은 아주 많이 나와있지만 이를 설문조사에 적용할 때 필요한 디자인 요소들에 대해 파악하여 제시함으로써 설문조사의 결과에 긍정적인 영향을 줄 수 있도록 도와주고자 한다.

## 제 4 절 연구의 방법

이러한 연구 목표를 달성하기 위해 본 연구에서는 Fluenty<sup>7</sup>와 Chatfuel<sup>8</sup>이라는 챗봇 제작 platform [48]을 통해 챗봇을 이용한 Rule Based Model[49]의 설문조사를 직접 제작하였다. 이렇게 만들어진 챗봇을 통한 설문조사를 실험하고 분석하는 등의 Iterative Design Process를 통해 몇 차례 개선하였다. 여러 번의 User Study를 통해 설문조사시 챗봇의 부드러운 말투나 감정의 교감 등의 Human Nature와 관련된 요소에 사용자들이 가장 영향을 많이 받음을 확인할 수 있었고 이에 대한 효과를 확인하기 위해 총 2가지 방식의 설문조사 - 온라인 설문조사, 친근한 챗봇 설문조사 - 를 가지고 최종 실험을 진행하도록 하였다. 설문 문항에 대해서는 개인 정보에 관한 질문, 객관적 사실을 묻는 질문 그리고 주관적 의견을 묻는 질문으로 나누어 구성하였으며 총 90여 문항이 각각의 설문지에 포함되었다. 실험 참가자는 IT기업으로부터 리크루팅 된 40명의 실험 참가자들이 각각 2주간 2가지 주제를 Chatbot과 Web을 통해 설문을 하도록 구성하였고 그 주제 및 방법은 무작위로 할당하여 설문을 진행하도록 하였으며 설문 결과의 밸런스를 위해 각각 20명씩 설문이 되도록 할당하였다. 설문을 통해 파악된 정량적인 요소들과 설문이 끝난 후 진행된 사후 인터뷰를

---

<sup>7</sup> Fluenty Inc.,. Fluenty.ai. <http://www.fluenty.co>

<sup>8</sup> Chatfuel. <https://chatfuel.com/>

통해 챗봇을 통한 설문조사 시스템의 유용성, 몰입도와 응답품질 및 사용자 선호도 등에 대한 자료를 수집하였다.

결과 분석을 위해서는 설문에 있어 중요한 요소인 응답품질에 대해 온라인 설문조사와 챗봇을 통한 설문조사를 비교 분석하였다. 기본적으로는 설문 문항의 종류인 개인정보, 객관적 사실 및 주관적 의견에 따른 응답품질을 각각 비교하여 분석하였으며 응답품질의 경우에는 객관식의 경우에는 응답의 성실도가 얼마나 좋은지 그리고 응답의 변별력이 어떻게 차이가 나는지 측정을 했으며 주관식은 응답이 얼마나 길고 무응답이 적은지에 대해 확인하였다. [33, 69] 추가적으로는 인터뷰에서 얻어진 몰입도[43], 심리적 거부감 및 유용성 [5]에 대한 분석을 진행하였다.

## 제 5 절 연구의 기여점

이전 연구에서는 온라인 설문조사와 챗봇 설문조사의 단순한 설문의 내용과는 상관없이 객관식, 주관식 문항의 답변 특징에 따른 응답 품질을 파악하여 유의한 차이를 보려고 하였지만 본 연구에서는 객관식, 주관식 문항의 기본적인 특징을 기반으로 하지만 질문의 내용을 한 단계 더 구조화 하여 주관적 의견, 객관적 사실 및 개인정보에 대한 질문으로 구분하여 유의한 차이를 파악하였고 이에 설문조사의 유형이나 목적에 따라 설문 조사 방법을 선택할 수 있는 학문적 기초를 마련했다고 볼 수 있다.

그래서 본 연구의 기여점은 1) 챗봇을 통한 설문조사에 대한 학문적 고찰을 진행한 연구에 해당하며 2) 챗봇이 기존 온라인 설문조사에 비해 개인정보, 주관적 의견 및 객관적 사실에 대한 응답 변별력, 응답 길이 및 무응답률에서 설문응답의 품질을 차이를 확인하여 설문조사의 종류에 따라 챗봇에 적합한 설문인지를 가늠해볼수도 있으며 몰입도, 유용성 및 거부감에 대해서 파악하였다. 그리고 실제 사용한 피실험자와의 사용자 연구와 평가 및 사후인터뷰를 통해 3) 설문조사를 위한 챗봇 제작에 있어 고려해야 할 사항들에 대해서 제안한 것이다.

## 제 2 장 선행연구

### 제 1 절 챗봇(Chatbot)

챗봇은 사용자와 자연어로 소통하는 대화형 에이전트이다. [8] 이러한 챗봇은 1966년에 발표한 ELIZA를 시초로 하여 [14] 처음에는 사람들로 하여금 실제 사람인 것처럼 인지하도록 하는데 초점을 맞추어 개발되었다. 이에 대한 판단을 위해 1950년대에 발표된 Turing Test를 이용하였으며 기계가 제시한 대화의 30%를 평가자가 사람으로 판단할 경우 통과하게 되는 기준으로 진행되었다. [76] 챗봇에서는 2014년 Eugene라고 명명된 컴퓨터 프로그램이 처음으로 Turing Test를 통과하게 되었으며 [77] 결국 이런 모든 노력들은 챗봇이 사용자의 의도를 이해하고 적절한 응답을 제공함으로써 실제 사람과 대화하는 듯한 감정을 사용자에게 느끼게 해주는 것이라고 볼 수 있다. 이를 위해서 NLP(자연언어처리), ML(기계학습) 및 DL(딥러닝)등과 같은 많은 신기술들이 챗봇에 적용하고자 하는 연구가 되고 있다. [12] 위에서 언급한 기술의 발전과 더불어 챗봇은 이미 다양한 산업 분야에서 널리 사용되고 있으며 대부분은 고객과 소통하고 제품 및 서비스를 판매하거나 정보를 검색하는데 사용이 되고 있다. 그리고 챗봇을 설문조사에 적용하는 시도도 없는 것은 아니며 챗봇을 통한 인터뷰를

통해 다양한 특징들을 발견하기도 하였다. NHN Entertainment에서 이뤄진 한 연구에서는 간단한 실험으로 대화형 챗봇이 사용자로부터 의견을 받는데 유용할 수도 있다는 의견을 제시하였다.<sup>9</sup> 그리고 서울대학교 융합과학기술대학원의 UX Lab에서 2017 HCIK에서 발표한 “챗봇으로 하는 인터뷰”<sup>10</sup>에서도 직접 설계한 챗봇 인터뷰를 통해 챗봇 설계시 고려해야 할 여러 가지 사항들에 대해 제안을 하기도 하였다. 하지만 위의 연구들은 챗봇으로 설문조사를 진행하는 실험을 통해 챗봇이 설문조사에 이용될 수 있다는 가능성을 정성적으로 확인하는데 그쳤으며 추가적인 실험이나 정량적 데이터 분석을 통해 어떠한 특징들이 챗봇을 통한 설문조사에 나타나지는 확인하지는 못하였다. 이에 이어진 연구로서 CHI 2019에 발표된 논문으로 실제 대화형 스타일이 설문조사 응답 품질에 어떤 영향을 주는지에 대한 연구[2]가 진행되었고 객관식을 통한 응답 변별력을 통해 응답 품질이 좋아진다는 연구결과를 선보였으며 대화형에 대한 실험 참가자의 다양한 의견을 수집하여 챗봇이 virtual interviewer로서의 가능성도 제시하였다. 하지만 기존 연구에서는 챗봇을 통한 설문조사에서 객관식에 대한 설문을 진행하여 실제 설문조사에 존재하는 다양한 항목들에 대한 실험이 부족했던 부분이 있다. 이에 본 연구에서는 가장 대표적인 설문조사 방법인 온라인 설문조사를 챗봇과 비교 실험하지만 객관식 및 주관식을 혼합하고 설문 조사에서 진행하는 개인정보 질문/ 객관적

---

<sup>9</sup> NHN Entertainment UI/UX Lab. 2017. [연구] 대화형 챗봇으로 사용자 리서치 하기. Naver Blog. [http://blog.naver.com/uiux\\_lab/221016593143](http://blog.naver.com/uiux_lab/221016593143)

<sup>10</sup> Jinyoung Kim et al. 2017. ‘Bot’ erview: Interviewing with Chatbot. HCIK 2017 Tutorial Session

사실/ 주관적 의견으로 질문형태를 나누어 이에 대한 정량적인 결과 데이터로 그 효과를 확인하고 추가적인 반구조화된 인터뷰를 진행하여 어떤 요소들이 챗봇에서 중요하게 다뤄져야 하는지를 파악해보고자 한다. 그리고 모든 결과를 종합하여 챗봇을 통한 설문조사를 디자인 할 때 고려해야 할 사항들에 대해서도 논의해보고자 한다.

## 제 2 절 다양한 설문조사 방법에 대한 연구

전통적으로 설문조사의 방법을 나열해보라고 하면 대면 설문조사, 전화 설문조사, 우편 설문조사 그리고 온라인 설문조사라고 할 수 있다. 이에 대한 장/단점은 이미 여러 연구에서 확인되었다.

Method of data collection	Advantages	Disadvantages
Personal	Complex questions Visual aids can be used Higher response rates	Expensive Time inefficient Training to avoid bias
Telephone	Allow clarification Larger radius than personal Less expensive or time consuming Higher response rates	No visual aids Difficult to develop rapport
Postal	Larger target Visual aids (although limited) Lower response rates	Non-response Time for data compilation
Electronic	Larger target Visual aids Quick response Quick data compilation	Lower response rates Non-response Not all subject accessible

표 1. 각 설문조사 방식의 장/단점 [82]

온라인 설문조사는 인터넷이 보급되기 시작하면서 먼저 이메일을 통한 설문조사가 가장 먼저 시작되었고 이후에는 웹을 기반으로 한 설문조사의 형태가 등장하게 되었다. 이메일로 회신을 요청하거나 ‘구글 Forms’ 와 같이 특정 웹사이트를 이용하기도 하며



혹은 ‘서베이몽키’ 처럼 멤버십을 이용하여 설문조사를 진행하는 방법도 모두 온라인 설문조사라고 할 수 있을 것이다. [50] Global Market Research(ESOMAR)에 따르면 전세계적으로는 2005년 설문조사 시장의 13%정도만 차지했던 온라인 설문조사가 2014년에는 26%까지 증가하였고 특히 미국, 일본 및 유럽의 선진국들은 이미 30~50%에 달하는 비중을 차지하고 있다고 한다.<sup>11</sup>

많은 수의 연구들이 온라인 설문조사의 효과성에 대해서는 기존의 설문조사 방법과의 비교를 통해 확인하였다. 전통적인 3가지의 설문조사 방법과의 비교 분석한 결과를 살펴보면 아래와 같다.

가장 오래되고 기본적인 방법으로 인터뷰어가 오프라인에서 대면으로 설문조사를 하는 방식이다. 이러한 대면조사와의 비교를 통해 타당도와 신뢰도를 기준으로 데이터 품질을 분석을 해보면 온라인 조사가 오프라인 대면조사에 비해 데이터 품질은 나쁘지 않고 대등한 수준을 보인다고 한다. [1] 그리고 온라인 설문조사가 훨씬 더 빠르고 비용도 적게 드는 반면 오프라인 대면조사가 훨씬 더 많은 정보를 수집할 수 있기도 하지만 [55] 어떤 경우에는 온라인 반구조화된 설문조사가 오히려 정보를 더 수집할 수 있다고 한다. [56] 하지만 대면조사의 경우 인터뷰어에 의한 편향이 발생하거나 [58, 59] 지역적인 한계가 있고 대량의 조사가 불가능하며 높은 비용이 든다는 [60] 단점이 있다.

---

<sup>11</sup> 이현우, 온라인 리서치 기법과 사례. <http://dure.net/ebiz98111.html>

우편을 통한 설문조사는 대면 설문조사와는 달리 인터뷰어에 의한 편향은 발생하지 않지만 시간이 오래 걸린다는 단점이 있는 방식이다. 온라인 설문조사가 우편 설문조사에 비해 비용[45]와 노동력에 대한 강점을 가지고 있지만 유사한 응답품질을 가지고 [52] 응답률은 더 낮지만 응답 시간은 더 짧으며 Open-ended Question에 대해서는 응답을 하지 않은 항목이 더 많지만 응답의 길이는 더 길게 나타나기도 했다. [53, 54]

전화 설문조사의 경우 직접 전화로 설문조사가 이루어지기 때문에 응답시간이 가장 빠르기는 하지만 대면조사와 마찬가지로 인터뷰어에 의한 편향이 생길 수도 있게 된다. 온라인 설문조사에서는 미 응답 항목에 대해서 사용자에게 알려줌으로써 이에 대한 비율이 줄어들게 되고 (하지만 전화 설문조사에서는 “의견 없음”의 답변을 받아들일 수 밖에 없다) Open-ended Question에서 시간을 더 사용하고 좀 더 정확한 응답을 하였으며 객관식 답변의 차별성은 적게 나타났다. [36]

최근에 나타나기 시작한 것이 바로 챗봇을 이용한 설문조사에 대한 연구이다. 대화형 시스템이라는 장점을 통해 고객 응대 혹은 상담 등의 분야에 대한 연구는 많이 진행되었지만 챗봇을 설문조사에 이용하는 것은 최근에 대두되고 있다.

Text-based의 챗봇이 온라인 설문조사와 비교해 새로운 설문 방법으로 응답 품질이 좋아지는지에 대한 연구[2]가 있었으며 이는 객관식 설문을 통해 Differentiation, Usability 및 Enjoyment를 측정하여 챗봇을 통한 설문조사가 더 차별화된 답변을 하게 되고 응답

품질이 좋아진다는 것을 확인하였고 몰입에 의한 Enjoyment 또한 응답 품질을 높이는 요인으로 작용하고 있다고 하였다. 그리고 이 연구에서는 정성적인 분석을 통해 챗봇을 통한 Survey의 여러가지 특징을 확인하였다. 챗봇 설문은 해야할 일이 아닌 Social Interaction으로 여기게 되고 Casual한 말투를 통해 사용자에게 좀 더 친밀감을 전달하고 있다고 하였으며 새로운 챗봇 설문이라는 경험을 통해 설문에 몰입하게 될 수 있게 하였으며 챗봇 설문 자체가 온라인 보다는 모바일 디바이스에 적합한 수단임을 확인하였다. 그리고 이전의 연구가 객관식을 통한 응답 품질을 확인하였다면 주관식 설문조사를 챗봇을 통해 실험을 하여 챗봇을 통한 설문조사가 사용자로 하여금 좋은 응답 품질을 가지게 되고 더 설문에 집중을 하게 된다는 것을 확인하였다. [80] 이 외에도 여러 특정한 분야에서 챗봇을 이용한 응답 품질에 대한 연구도 이루어졌으며 Course Evaluation에 대한 설문조사에서도 챗봇을 이용할 경우 응답품질이 더 높아지고 사용자의 Enjoyment가 증가한다는 연구결과도 확인되었다. [79]

본 연구에서는 기존 연구 대비하여 객관식 설문이나 주관식 설문만을 각각 따로 실험한 것이 아니라 일반적인 설문의 형태인 객관식 및 주관식 설문을 병행하여 설문을 진행하여 객관식에서의 differentiation 뿐만 아니라 open-ended question에 대한 응답 길이 및 No Response를 측정하여 다양한 관점에서 응답 품질을 확인하고자 하였다. 그리고 이전 연구에서는 설문의 내용과는 무관하게 객관식, 주관식 문항에서의 응답 품질의 차이만 보려고 하였지만 본 연구에서는 객관식, 주관식 문항의 기본적인 특징을 기반으로 하되 질문의 내용을

한 단계 더 구조화 하여 주관적 의견, 객관적 사실 및 개인정보에 대한 질문으로 구분하여 유의한 차이를 파악하고자 하였으며 기존 실험들은 설문지 특징에 기반하기 보다는 일반적인 설문에서 그 응답품질을 확인하였기 때문에 어떤 설문인가에 따라 그 효과가 나타날 수 있을지를 일반화 하지는 못하고 있다. 이에 본 연구에서는 설문지 특징에 따른 챗봇 설문조사의 효과를 파악하여 유의미한 차이가 있는지를 보려고 하였다.

실험 참가자와의 사후 인터뷰를 통해서도 실제로 사용자가 느끼는 몰입도나 유용성에 대해서 정량적으로 온라인 설문조사와 비교하였으며 사용자의 인터뷰를 기반으로 챗봇 설문지 대한 전반적인 사용자 경험에 대해 논의를 하였으며 챗봇 설문조사에 대한 앞으로 챗봇으로 설문 조사를 디자인할 때 고려해야 할 사항들에 대해 제안을 함으로써 추후 챗봇 설문 조사를 연구하거나 실제 챗봇 설문을 디자인할 때 참고할 수 있도록 하였다.

### 제 3 절 설문조사의 평가 방법

설문조사의 평가를 위해서 명확한 이론적인 방법이 정해져 있는 것은 아니지만 대부분의 연구자들은 응답률[31, 61, 62]과 응답품질[63]에 대해 측정하여 분석하고 있다. 특히 응답률이 낮으면 Nonresponse Bias[13, 65, 66]나 Nonresponse Error[64]의 가능성이 높아지게 되며 Nonresponse Error는 샘플의 많은 수가 설문에 응답하지 않아 전체 결과에 영향을 미치게 되는 경우를 말하며 Nonresponse Bias는 응답자의 조사 결과가 비 응답자의 것과 다를 경우를 말한다.

낮은 응답률을 가질 경우 Sample의 크기가 줄어들기 때문에 통계적인 약점으로 작용할 수 있으면 Nonresponse Bias의 문제도 생기기 되어 설문조사 결과에 대한 유효성에도 영향을 줄 수도 있게 된다. 이처럼 설문조사를 진행함에 있어 중요한 문제로 인식이 되고 이에 대한 개선을 위한 연구들이 존재한다. 응답률을 높이기 위해 인센티브를 주는 방식이 가장 흔하게 나타나며 비용을 높이는 원인이 되기도 하지만 가장 효과적인 방법이기도 하다. [34] 질문의 길이가 짧을수록 그 응답률을 더 길어지는 것을 확인하기도 하였고 [68] 얼마나 반복적으로 설문을 요청하고 사후 관리가 되었는지 그리고 조사주제도 응답률에 직접적인 관련이 있다고 보고 있다. [28] 위와 같이 언급한 것처럼 응답률을 높이기 위해서 해당 설문방법 내에서 그

해답을 찾으려고 노력하고 있다. 하지만 이 논문에서는 챗봇이라는 새로운 방법을 적용하여 대화형 인터페이스가 어떻게 응답률에 기여할 수 있는지를 온라인 설문조사와 비교하여 확인 해 보고자 한다.

전통적으로는 응답 데이터의 품질은 Close-ended나 Open-ended Question에서 응답하지 않는 항목의 양이나 Open-ended Question에서 단어의 수로 측정되는 응답의 길이를 기준으로 삼았다. 왜냐하면, 보통은 좀 더 긴 답변이 그 응답의 품질 면에서 더 낫다고 판단을 하였기 때문이다. [53] 그리고 응답 시간이나 Grid Question에 대한 답변의 다양성도 응답품질에 관련이 있다고 연구되었다. [33] 흔히 이러한 응답품질의 지표를 좋게 하기 위해 질문의 수를 줄이거나 마찬가지로 인센티브를 주는 등의 방법이 확인되었다. [33, 7] 이 논문에서는 새로운 챗봇이라는 도구를 사용하여 설문조사를 진행함으로써 응답품질에는 어떤 변화가 있는지를 온라인 설문조사와 비교하도록 하고 챗봇에서도 어떠한 요소를 강조해야 더 긍정적인 효과가 나타나는지도 확인해보고자 한다.

본 연구는 챗봇이라는 새로운 도구로서 온라인 설문조사와의 비교 실험을 통해 설문조사를 평가하는 가장 기본적인 척도인 응답품질을 응답 길이, No Response (DK Response) 및 Differentiation [69]을 통해 정량적 분석을 해보고자 한다.

## 제 3 장 연구 문제

### 제 1 절 연구 문제

이 절에서는 챗봇을 통한 설문조사의 실험을 통해 온라인 설문조사와의 차이점을 파악하고 친근한 말투의 챗봇 설문조사가 응답 품질에 어떤 영향을 미치는지를 찾는 것을 연구 문제를 설정하고, 그 효과가 개인정보, 객관적 사실 및 주관적 의견과 같은 설문의 종류에 따라 어떤 차이를 보이는지에 대해 정량적인 데이터의 분석을 기반으로 그 관계를 설명하고자 한다. 선행연구에서도 확인했듯이 설문조사에서는 응답률과 응답의 품질이 중요하며 이는 주관식에서의 응답 길이나 객관식에서의 변별력 등을 기준으로 판단하게 된다. 또한 정성적인 요소로서 설문에 대한 심리적인 거부감, 몰입감, 유용성 등도 파악하여 새로운 설문조사의 방법으로서 유용한지도 확인해보고자 한다.

[연구 문제 1] 온라인 및 챗봇 설문조사의 비교를 통해 어떠한 차이점을 발견할 수 있는가?

H1. 메신저를 기반으로 하는 챗봇 설문조사는 주관적 의견 설문에 대한

주관식의 응답길이가 더 길고 (H1a) No Response도 적을 (H1b) 것이다. 객관식에 대한 설문에 대한 응답 Differentiation이 크고 (H1c) No Response도 적을 (H1d) 것이다.

H2. 메신저를 기반으로 하는 챗봇 설문조사는 객관적 사실 설문에 대한 주관식의 응답길이가 더 길고 (H2a) No Response도 적을 (H2b) 것이다. 객관식에 대한 설문에 대한 응답 Differentiation이 크고 (H2c) No Response도 적을 (H1d) 것이다.

H3. 메신저를 기반으로 하는 챗봇 설문조사는 개인정보에 대한 무응답률이 더 적을 것이다.

H4. 메신저를 기반으로 하는 챗봇 설문조사의 사용자 몰입도가 더 좋을 것이다.

H5. 메신저를 기반으로 하는 챗봇 설문조사의 유용성이 더 좋을 것이다.

연구 문제 H1에서 H3은 설문조사 테스트를 통한 정량적인 데이터를 기반으로 유효한 차이점이 나타나는지를 확인하고자 한다. 이는 챗봇이 설문조사에 이용됨으로써 얼마나 효과적인지를 파악하고 추후 설문조사를 위한 일반적인 방법이 되기 위한 이론적인 기반이 될 수 있는지 확인해보고자 한다. H4, H5는 피실험자와의 인터뷰를 통한 정성적인 피드백의 분석을 통해 챗봇을 통한 설문조사의 몰입도와 유용성에 대해 파악하고자 한다.



[연구 문제 2] 챗봇을 통한 설문조사를 제작할 때 어떠한 요소들을 고민해야 하는가?

연구 문제 2에서는 앞서 파악된 정량적이고 정성적인 결과를 기반으로 챗봇을 통한 설문조사에 대한 실증적인 이해와 더불어 어떠한 시사점을 가지는지를 알아보도록 한다. 그리고 인터뷰를 통한 사용자들의 다양한 의견을 확인하고 그 상관관계를 파악하여 챗봇을 설문조사라는 것에 적용하고자 할 때에는 어떠한 고려사항들이 있는지를 도출할 수 있을지에 대해 알아보고자 한다.

## 제 4 장 연구 방법

### 제 1 절 연구방법 요약

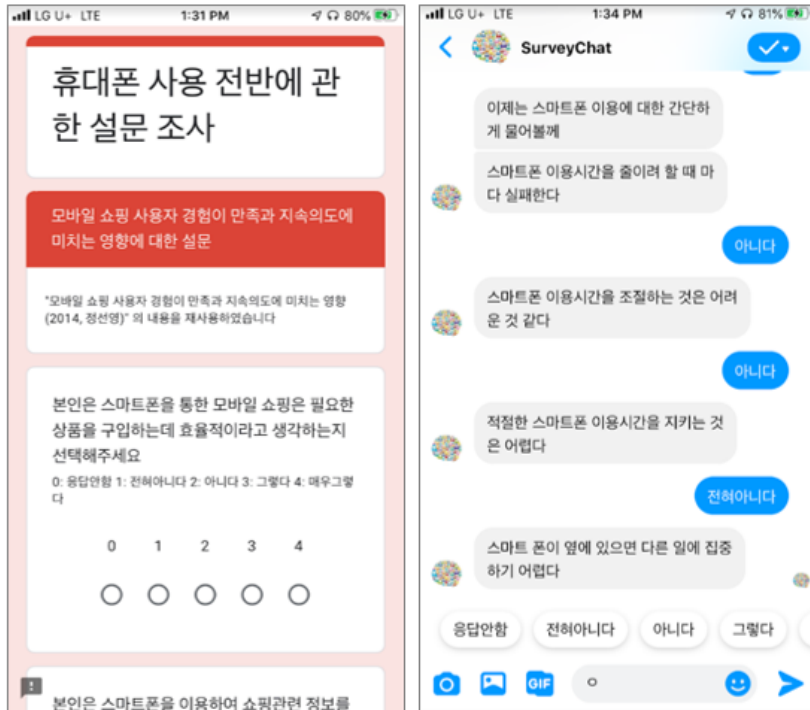
이 연구의 목적은 온라인 설문조사와 챗봇을 이용한 설문조사를 비교 실험 및 분석을 하여 응답품질이 어떻게 다른지에 대해 확인해보고 챗봇을 이용한 설문조사를 디자인할 때 고려해야 할 사항에 대해 파악해보는 것이다. 이에 최종 실험을 위한 챗봇을 디자인하고 실험을 진행하기 위해 Iterative Design Process를 통해 여러 회에 걸쳐 챗봇을 통한 설문조사를 디자인하고 실험을 진행하여 피드백을 통해 중요한 요소를 찾아내도록 하였다. 이 과정을 통해 챗봇이라는 대화형 시스템에서는 사용자가 느끼기에 부드러운 말투나 감정의 교감 등의 Human Nature와 관련된 요소에 가장 큰 영향을 받았음을 확인하였고 이를 토대로 최종 실험에서는 개인정보, 객관적 사실 및 주관적 의견의 설문에 대한 온라인 설문조사와의 차이점에 대해 파악하기 위해 아래 2가지 type의 설문조사를 진행하기로 하였다.

- Online Survey (온라인 설문조사): 가장 일반적인 ‘Google Forms’를 이용하여 제작하였으며 실험 참가자의 메신저를 통해 질문이 전달되도록 하였다. 질문의 내용은 일반적인

설문조사에서 사용하는 문어체로 작성되었다.

- Intimate Chatbot Survey (친근한 챗봇 설문조사): 챗봇을 사용하여 설문조사를 진행하며 온라인 설문조사나 챗봇 설문조사와는 달리 일반적으로 메신저에서 주고 받는 구어체로 질문내용은 작성되었다. 챗봇은 Chatfuel Platform을 이용하여 제작되었으며 페이스북 메신저상에서 동작한다.

설문 주제는 2가지로 준비되었으며 각각의 주제는 스마트폰 과의존 실태 설문조사 (설문주제1) 와 휴대폰 사용 전반에 대한 설문조사 (설문주제2) 로 작성되었으며 실험참가자가 특별한 지식이 없어도 쉽게 답변할 수 있는 항목으로 구성되어 있다. 설문주제1은 과학기술정보통신부가 주관한 “2018년 스마트폰 과의존 실태 조사의 설문”을 인용하여 재구성하였고, 설문주제2는 “모바일 쇼핑 사용자 경험이 만족과 지속의도에 미치는 영향(2014, 정선영)”의 설문조사 내용을 재구성하였다.



(a) Web Survey

(b) Chatbot Survey

그림 1. 2가지 다른 설문조사 방법에 대한 예시

위와 같은 2가지의 설문주제를 가지고 챗봇과 온라인 설문조사와의 비교를 위해 40명의 피실험자를 모집한 후 실험을 진행하였고 반 구조화된 사후인터뷰를 진행하여 어떤 특징들을 확인할 수 있는지를 파악하고 챗봇을 이용한 설문조사에서 나타나는 여러가지 결과 및 특징들을 분석하도록 하였다. 본 연구에서는 챗봇의 인터페이스 자체에 대한 효과를 분석하기 위해 Rule-Based Model로 작성되어 있으며 실험 참가자의 응답에 따라 Interactive하게 설문 항목이 바뀌거나 그 설문의 문항 개수가 바뀌지는 않도록 디자인 하였다.

## 제 2 절 User Study #1 (챗봇 설문문의 인식 조사)

우선 챗봇을 통한 설문조사가 많이 사용되고 있지 않기 때문에 이 연구에 대한 가능성을 파악하고자 하였다. 챗봇 설문조사에 대한 사용자의 Impression을 파악하기 위해 챗봇을 이용해 설문조사를 진행하는 예제를 만들었다. 이는 Fluety.ai platform을 이용한 Facebook Messenger를 기반으로 동작을 하며 서울대학교 학습개발센터의 강의 후 강의에 대한 평가와 유사한 형태로 만들어졌다. 저자가 직접 설문조사를 진행하고 동영상으로 캡처하여 피실험자에게 배포하였고 이를 통해 다양한 의견을 들어보게 되었다. 본 실험에서 받은 여러 피드백을 기반으로 연구에 대한 필요성을 확인하였으며 사후 인터뷰를 통해 챗봇의 장단점에 대한 의견도 확인하여 이를 최종 실험에

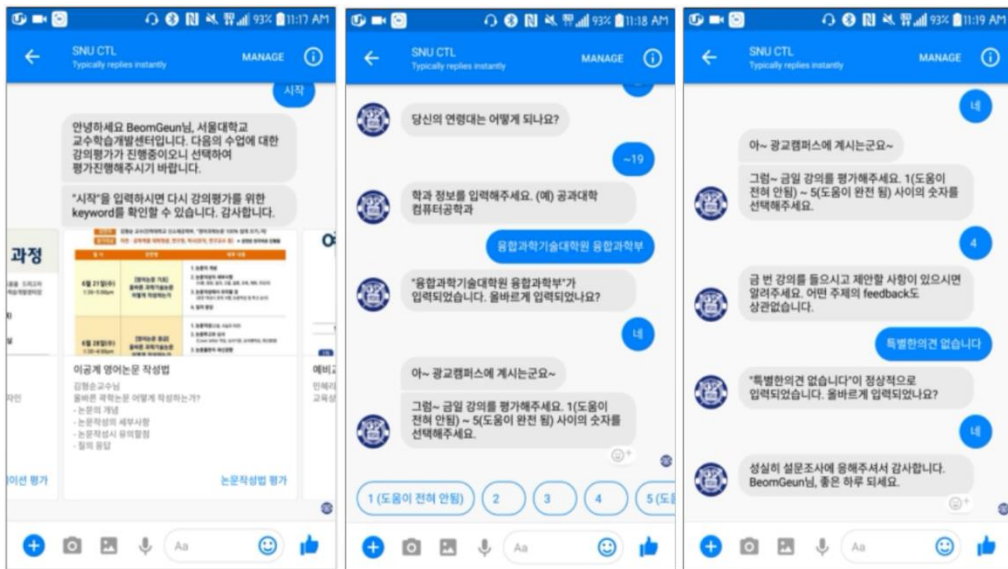


그림 2. 챗봇을 통한 설문조사의 예시 동영상

사용할 챗봇 설문조사 디자인에 반영하였다.

이 설문조사에서는 챗봇을 통한 설문조사를 접해보지 못한 일반인이 새로운 설문조사 방식으로서 챗봇을 보았을 때 어떻게 인식될지에 대한 기초적인 질문에서 시작되었으며 이를 통해 앞으로의 연구가 가능성이 있는지에 대해 탐색해보고자 하였다. 그래서 Fluently.ai platform을 이용하여 강의 후 평가에 대한 10문항의 설문조사를 Chatbot으로 만들어 이를 13명(Mean=31.77, SD=2.98; 남성 9명, 여성 4명)의 IT 기업 종사자들에게 챗봇 동영상상을 메일로 공유한 후 이에 대한 응답률과 응답품질에 대한 각자의 의견을 답변 받도록 하였다. 챗봇을 사용했을 때 응답률과 응답품질이 기존 IT기업 내에서 진행되는 온라인 설문조사에 대비하여 얼마나 개선이 될지에

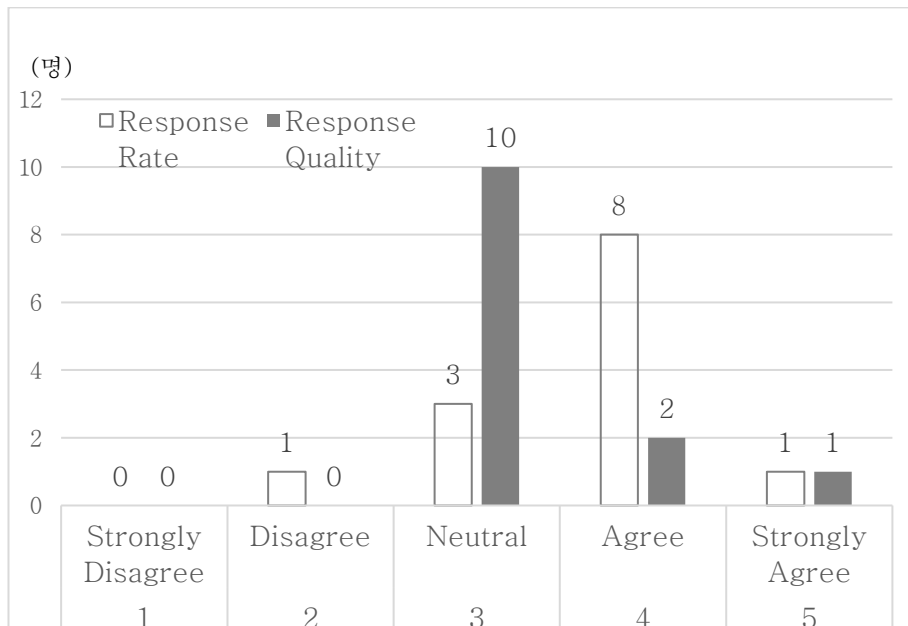


그림 3. 챗봇이 온라인 설문조사보다 개선될 것인가에 대한 설문 결과

대해 5-point Likert Scale Rating에서 (1) Strongly Disagree에서 Strongly Agree(5)로 측정하였다.

응답품질에 대해서는 좋아진다는 의견보다는 기존 온라인 설문조사와 비슷할 것이라는 의견이 압도적이었으며 응답률의 경우에는 좋아질 것이라는 의견이 대체적으로 많았다. 이와 같이 어느 정도 긍정적인 면의 가능성을 보았기 때문에 이 연구를 계속 수행할 수 있었고 이후 실제 사용자들이 경험해볼 수 있는 ‘ChatSurvey’를 만들어 User Study를 진행해 나갔다.

1차 설문조사의 사후 인터뷰에서 자신이 해당 점수를 준 이유에 대해서 확인하였고 그 외의 챗봇 설문조사에 대한 피드백을 받아보았다. 응답률에 대해서 긍정적인 답변을 한 실험 참가자들은 대체적으로 챗봇의 접근 용이성을 강점으로 생각하였다. 메신저를 기반으로 하고 있기 때문에 사람들이 언제 어디서나 확인가능 하고 자주 사용하면서 사용 시간도 가장 긴 앱에 속하기 때문으로 판단하였다. 응답률에 대해 유일하게 Worse라고 체크한 실험자는 아래와 같이 얘기하였다.

*(P13) 챗봇의 UI가 그렇게 편한 것 같지는 않다. 아무래도 모바일에서 사용하는 키보드의 한계도 있고 답변을 기다려야 하는 등 속도가 느려 보여서 바쁜 사람이라면 온라인 설문조사가 더 선호되지 않을까 싶다.*

응답 품질에 대해서는 품질이라는 것은 대부분 개인의 성향차이가 있기 때문에 전체적으로 보면 크게 다르지 않을 것이라고 대부분의 실험자가 예상을 하였다. 다만 대화형식이 부드럽게 때문에 응답품질이 조금은 더 나을 것이라고 판단하는 실험참가자도 있었다.

(P3) 메신저를 통해 대화형식으로 사용되기 때문에 우선 부드러운 말투로 대화하듯이 설문을 하게 된다면 응답의 품질도 높아질 것으로 예상이 된다.

그 외 챗봇을 사용하면서 느낀 부분에 대한 실험자들의 의견은 개인 메신저를 설문조사에 이용되는 것이 Privacy에 대한 불편함을 느낄 수도 있다라는 의견(P5, P6, P7, P8, P9)이 많이 있었다. 그리고 이전의 내용을 확인하거나 수정하기가 어려울 것 같다는 실험자(P3)도 있었으며 Progress를 알 수가 없고 만약 긴 설문일 경우 온라인에서 PC에서 하는 것이 나을 것 같다(P12)도 있었다.

정리하자면, 실제로 챗봇으로 설문조사를 진행하지는 않았지만 챗봇을 통한 설문조사의 간단한 동영상 확인한 13명의 실험자들은 응답률은 올라갈 가능성이 있고 응답품질은 비슷할 것이라는 예상을 했다. 그 외에 Privacy나 모바일 설문조사에서 나올 수 있는 단점을 지적하였다. 이런 실험자들의 예상에 근거하여 챗봇을 통한 설문조사의 가능성을 확인하였고 이를 바탕으로 실제 ‘ChatSurvey’를 제작해서 피실험자들이 사용해볼 수 있는 User Study를 진행하게 되었다.



### 제 3 절 User Study #2 (챗봇 설문조사 실험)

이 실험에서는 실제로 실험자들이 사용해볼 수 있는 “ChatSurvey”를 만들어 실제 경험에서 나오는 피드백과 챗봇 설문조사에 대한 의견을 듣고 이를 최종 실험을 위한 챗봇을 제작하는데 있어 참고를 하고자 하였다. 실험자는 총 8명(Mean=29.13, SD=3.14; Male 6명, Female 2명)이며 대학원생을 대상으로 하였다. 실험 참가자 중에서는 챗봇에 대해 다 알고 있었고 써본 사람이 4명이었고 그 중 챗봇을 만들어 본 사람도 1명이 포함되었다. 본 User Study에서도 역시 Fluently.ai platform을 그대로 사용하고 Rule Based Model의 Facebook Messenger에서 실행이 되는 “ChatSurvey”를 통해 카카오뱅크와 같은 신규 비대면 인터넷 은행에 관련한 설문을 진행하였다. 챗봇 실험 중에는 Think-aloud Method[71]를 이용하여 기록하였고 실험 후 모든 실험 참가자로부터 반 구조화 된 인터뷰[72]를 진행 하였다. 실험에서는 참가자들이 피로를 느끼지 않게 15개의 질문을 10분이 걸리지 않도록 구성<sup>12</sup>[73]하였고 보통 참가자들은 3분 ~ 5분의 시간이 걸렸다.

---

<sup>12</sup> How Many Questions Should I Use in My Survey? (Surveygizmo)  
<https://www.surveygizmo.com/resources/blog/how-many-survey-questions/>

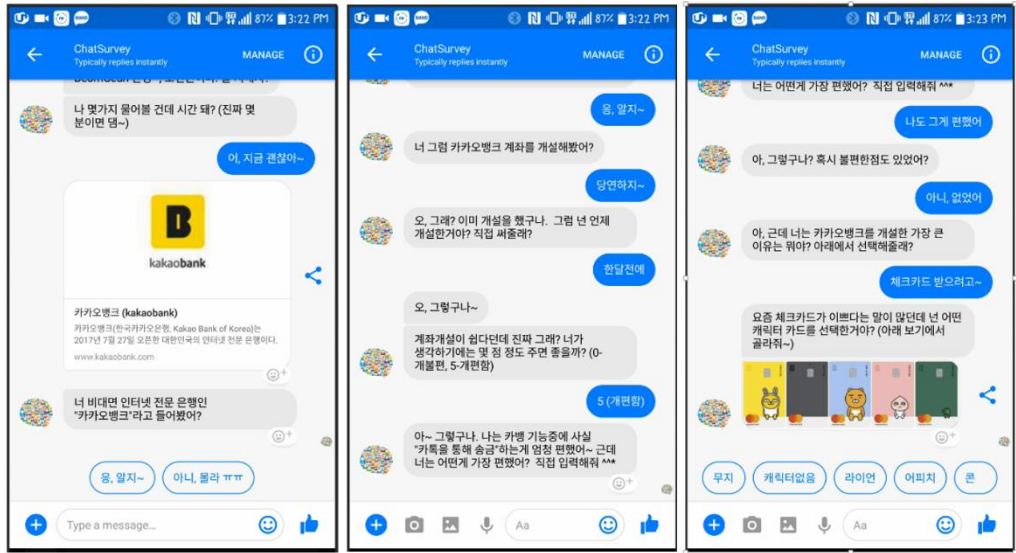


그림 4. 카카오프뱅크 설문에 대한 챗봇 화면

Think-aloud method와 사후 인터뷰에서 나온 결과를 Thematic coding method[75, 74]를 이용해서 분석하였으며 종합해보면 다음과 같다. 우선, 사용 중에 나온 피드백에서는 챗봇을 이용한 설문조사의 장점으로 메시지를 이용한 대화형식이기 때문에 몰입이 되며 말투가 친근해서 설문조사 하는데 거부감이 적게 든다는 것이었다. 단점으로는 전체 progress를 확인할 수가 없었고 혹시라도 잘못된 대화형 어투를 사용할 경우 사용자에게 불쾌감을 줄 수도 있다는 부분도 지적해주었다.

(P1) 온라인 서베이는 잘 하지 않는 데 이 챗봇 설문조사는 다음에 무슨 문제가 나올지 궁금해진다. 그래서 재미있다. (P3) 말투자체에서 오는 재미가 있다. 대화형식이기 때문에 거부하고 싶은 느낌은 없고 다음에 무슨 얘기를 할까 궁금해지기도 한다. 기존 설문조사의 경우에는

뭔가 채워야 할 숙제 같은 느낌이었다. Chatbot Survey는 스토리텔링이 될 것 같다. (P4) 대화형임에도 불구하고 설문을 한다는 느낌은 들지만 사람이랑 하는 것 같더라는 느낌이 드는데 이건 대화형, 특히 말투 때문인 것 같다. 캐주얼해서 설문에 대한 장벽이 낮아질 것 같다. (P6) 채팅 하듯이 하니까 일하거나 억지로 하는 느낌보다는 편한 느낌이 있다. 채팅이다 보니 구어체로 자연스럽게 하게 되는 것 같다. (P7) 말투 같은 것이 휴먼라이크하다. 무미건조하면 재미가 없다. 다 코딩을 해 놓았기 때문에 잘 알고 얘기하는 것 같다. 설문에서 빠져나갈 생각이 들지 않게 몰입감도 괜찮았다. 일방적인 의견을 끌어 내는 것이 아니라 자신의 의견을 suggestion<sup>13</sup> 하니까 좋은 것 같다.

이 외에는 아래와 같은 의견이 추가로 있었다.

(P2) 말투 때문에라도 대화형의 장점이 돋보인다. 하지만 말투에 대해서는 variation이 큰데 이에 대한 design guide도 필요할 것 같다. (P6) 챗봇이라는 걸 이해하게 되면 혹시라도 알아듣지 못할까 봐 이해시키려고 노력하게 될 것 같다. (P7) 피 설문자가 빠져나가지 못하도록 디자인을 해야 할 것 같다. 단발성 의견수집에는 적합할 것 같지만 정확한 의견을 복수로 모아서 수집해야 하는 경우는 기존 설문도 나쁘지 않다고 생각된다. (P5) 대화형이라 주의는 끌 수 있을 것 같으나 학습효과에 대해서는 고려해봐야 할 것 같다. (P8) 기업에서 챗봇을 통해 소비자 피드백을 얻는 데 있어 장벽을 낮춰줄 수 있을 것이라

---

<sup>13</sup> 예시를 들어주고 묻는 형식으로 표현한 것이 있었다. (예시) “나는 카뱅기능 중에 사실 ‘카톡을 통해 송금’ 하는 게 엄청 편했어~ 근데 너는 어떤 게 가장 편했어?”

생각한다, 그러나 장벽이 낮아진 만큼 *gaming behavior*를 방지할 수 있는 유인 또한 제공/디자인해야 할 것 같다.

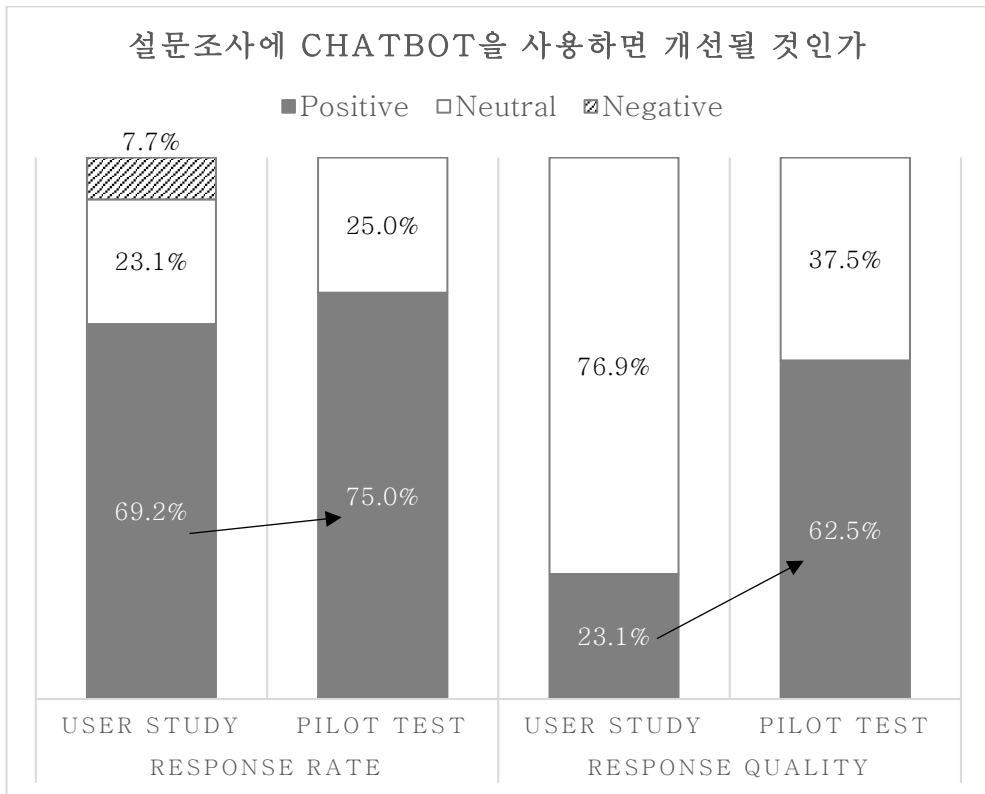


그림 5. 챗봇 설문 사용 후 응답률 및 응답품질 예측의 변화

그리고 응답률에 대해서는 50%인 4명은 온라인 설문조사에 비해 비슷할 것이라는 예상을 하였고 나머지는 좀 더 나아질 것이라는 예상을 했다. 응답품질에 대해서는 주관식문제에 대해서는 모바일 키보드 및 대화형의 경험으로 인해 단답형 답이 많이 나올 것 같으며 반면 객관식에 대해서는 대화형이어서 좀 더 심사숙고를 하게 된다는 의견이 있었다. 이에 본 실험에서는 모바일을 통한 설문을 진행하도록 요청하여 입력 수단의 차이로 인한 결과의 차이를 없애기로 했다.

실제로 사용을 해본 후 응답률과 응답품질에 대해 기존 User Study와 비교를 해보면 아래와 같이 응답률과 응답품질 모두 상승하였으며 응답품질에 대해서는 좀 더 큰 변화가 감지가 되었다. 이는 실제로 사용해 봄으로써 챗봇을 이용한 설문조사가 편하면서 몰입감을 주고 있다는 것을 말해주는 것이다.

(P7) 응답률은 높을 것 같습니다. 진입장벽이 살짝 낮고 일단 챗봇이라는 형태 자체가 일반 경험보다는 특이하기 때문에 적어도 호기심에서라도 시작하는 사람이 플러스 될 것 같습니다. 그리고 온라인 서베이라고 하면 예상되는 거부감을 상쇄할 수도 있을 것 같습니다. (P8) 편의성 면에서는 당연히 챗봇을 더 선호할거 같다. ... 응답률이 높아지니 측정치는 많아질 수도 있고, 빠른 응답률을 기대해볼 수 있으나, 주관식 답변에 있어서는 유저의 성향에 따라 다르겠지만 온라인보다는 짧을 수 있을 것 같다. 성의 없는 행동을 줄일 수 있도록 하면 no answer도 줄어들 수 있지 않을까? (P7) 근데 몇 가지 단점도 있을 것 같습니다. 예를 들어 질문들 간의 상대적인 요소를 파악 하는 데는 어려움이 있을 것 같습니다. 질문이 진행되는 맥락을 놓칠 수도 있을 것 같습니다. 그리고 주관식 응답 길이의 경우 좀 단편적으로 쓸 것 같습니다. ... No answer는 아무래도 좀 줄어들지 않을까 싶습니다. 대화 중에 '없어' '몰라' 이렇게 무시하는 것이 좀 덜할 것 같아요. 무엇 하나라도 의견 몇 개를 덧붙이게 될 것 같습니다. (P2) Chatbot Survey의 경우에는 메신저를 기반으로 하고 있기 때문에 중간에 중단을

하더라도 아무 때나 다시 돌아올 가능성이 큰 것 같다. (P4) 체크하고 채점하는게 아니라 대화형식이 주는 특징으로 관심 있는 주제라면 사람과 얘기하듯이 진행하면서 훨씬 좋은 feedback을 받을 수 있는 수단이 될 것 같다. (P7) 단발성으로든 일단 사용자의 의견을 빠르게 캐치할 수 있다는 장점이 있지만, 그걸 깊이 있게 알아내려면 답변을 이해하고 추가 질문으로 이어져야 하는데 그걸 못한다면 오히려 독이 될 수 있을 것 같습니다. (P3) 대화 중에 suggestion을 하는데 이는 답변을 유도하는 느낌이 있다. 그래서 자칫하면 객관성을 잃어 버릴 수 있게 될 것 같다.

결과를 요약해보면 “ChatSurvey”는 대화형에 의한 몰입, 말투에서 나오는 친근하고 편안함 그리고 메신저라는 연속적인 경험에서 나오는 낮은 거부감이 있으며 기업에서 소비자의 반응에 대해 빠르고 간단하게 설문 결과를 얻기에 좋을 것 같으며 이와 더불어 사용자가 느낄 수 있는 지루함이나 말투 사용의 문제 및 객관성 유지를 위한 유연한 디자인이 필요하다는 조언이 있었다.

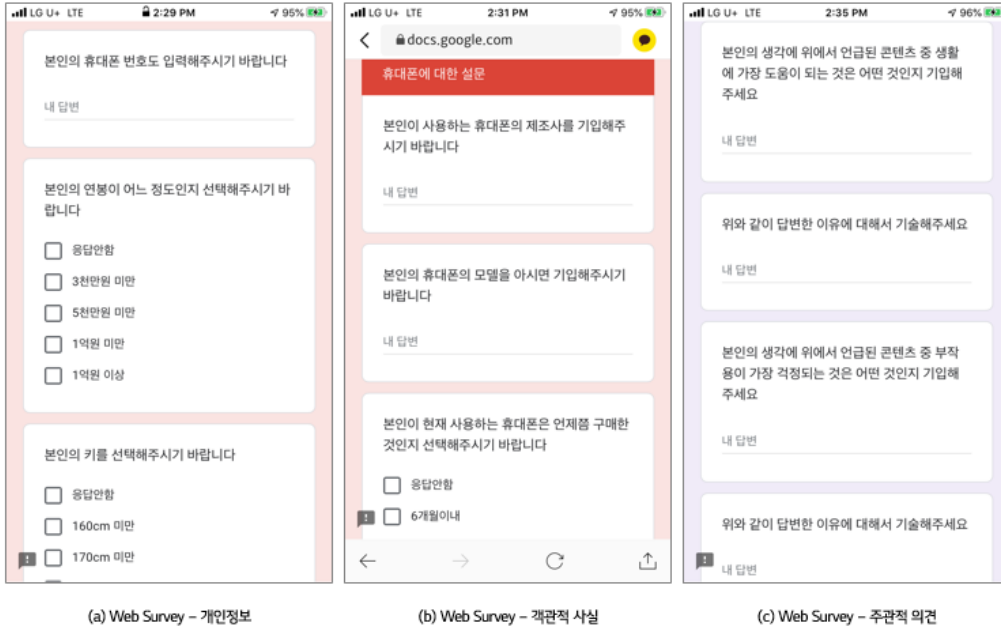
## 제 4 절 Evaluation

이 연구에서는 친근한 말투의 챗봇을 통한 설문조사가 온라인 설문조사에 비해 어떤 특징을 갖고 응답률이나 응답품질이 개선 될 수 있는지에 대해 알아보기 위하여 여러 번의 User Study의 결과를 바탕으로 2가지의 설문조사 형식(온라인 설문조사, 챗봇 설문조사)에 대해 개인정보, 객관적 사실 및 주관적 의견에 대한 설문항목을 설정하였고 실제 설문조사 실험을 위한 Topic을 정의하고 실험참가자들이 2가지 형식에 대해 서로 다른 설문주제를 가지고 설문을 진행하도록 하여 그 결과를 비교 분석하도록 하였다. 실험이 완료된 후에는 반구조화된 사후 인터뷰도 진행하였으며 모든 인터뷰의 내용은 저자에 의해 전사가 되었다.

### 설문 디자인

이전의 User Study의 실험 결과에 따라 기본적인 챗봇을 이용한 설문조사를 구현함에 있어 말투와 사용자와의 반응으로 통해 의인화의 요소를 가미한 Chatbot Survey로 설문을 진행하였다.

- 온라인 설문조사 (온라인 설문조사): 가장 일반적인 Google Form을 이용하여 제작하였으며 실험자의 Facebook Messenger를 통해 설문 링크는 전달되었다. 질문의 내용은 일반적인 설문조사에서 사용하는 문어체로 작성되었다.



## 그림 6. 온라인 설문조사 항목

- Intimate Chatbot Survey (친근한 챗봇 설문조사): 챗봇을 사용하여 설문조사를 진행하며 온라인 설문조사와는 다르게 대화형 인터페이스에서 자주 사용되는 구어체로 그 내용은 작성되었다. 챗봇은 Chatfuel Platform을 이용하여 제작되었으며 페이스북 메신저상에서 동작하며 설문 요청은 메신저로 직접 요청되었다.



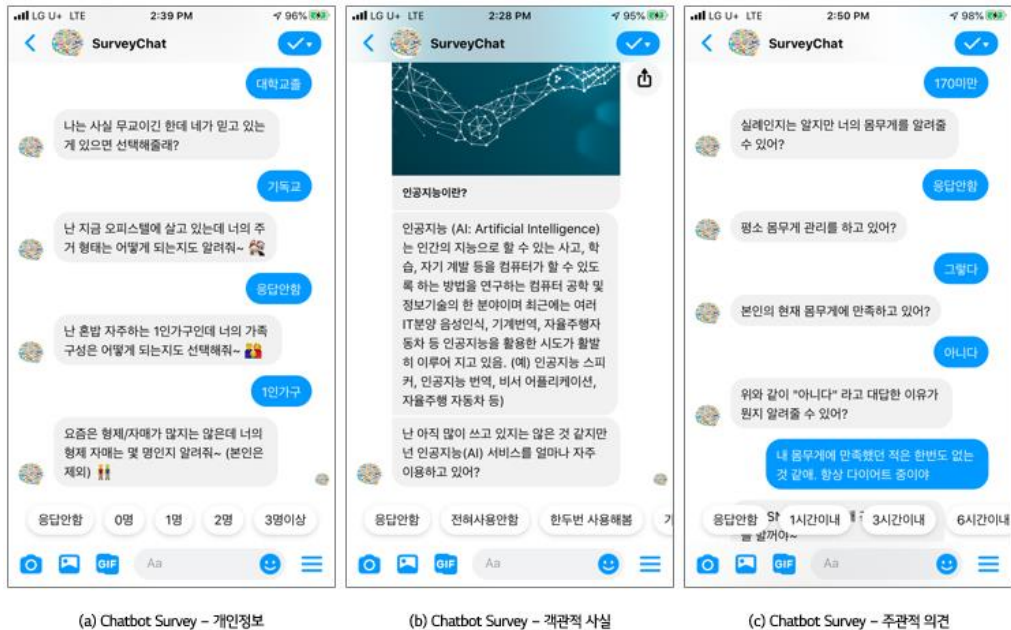


그림 7. 챗봇 설문조사 항목

설문 주제는 2가지로 준비되었으며 각각의 주제는 스마트폰 과의존 실태 설문조사 (설문주제1) 와 휴대폰 사용 전반에 대한 설문조사 (설문주제2) 로 작성되었으며 실험참가자가 특별한 지식이 없어도 쉽게 답변할 수 있는 항목으로 구성되어 있다. 설문주제1은 과학기술정보통신부가 주관한 “2018년 스마트폰 과의존 실태 조사의 설문”을 인용하여 재구성하였고, 설문주제2는 “모바일 쇼핑 사용자 경험이 만족과 지속의도에 미치는 영향(2014, 정선영)”의 설문조사 내용을 재구성하였다. 설문주제 1의 경우 총 93문항으로 개인정보 10문항(객관식 5문, 주관식 5문), 객관적 사실 41문항 (객관식 37문, 주관식 4문) 그리고 주관적 의견 42문항 (객관식 32문, 주관식 10문)으로 구성되어 있고 설문주제 2의 경우 총 89문항으로 개인정보

9문항 (객관식 3문, 주관식 6문), 객관적 사실 47문항 (객관식 39문항, 주관식 8문항) 그리고 주관적 의견 33문항 (객관식 8문항, 주관식 25문항)으로 구성되어 있다. 각 설문조사는 10분 ~ 20분 정도 소요되는 것으로 나타났다.

설문을 위한 재구성은 객관적인 사실에 대한 설문 항목은 많으나 주관적인 의견에 대한 설문항목이 많지 않았기 때문에 주관적 의견에 대해 질문 할 수 있도록 주로 객관식 항목을 선택한 이유 등을 질문함으로써 이를 보완하였다.

Survey Types를 실험하기 위한 content의 경우 실험참가자의 차이에 의해 편향이 발생하지 않도록 전문적이거나 깊은 지식이 필요한 내용 대신 일반인이 누구나 쉽게 대답할 수 있는 주제를 가지고 100개 내외의 문항이 20분이내에 완료될 수 있도록 선정하였다.

실험 참가자들이 진행할 설문조사는 아래와 같이 총 4개(2 Types X 2 Topics)가 준비되어 있으며 설문 주제에 따라 각각 총 93문항, 89문항의 설문항목들로 구성되어 있다. 객관식 문항의 경우 응답의 변별력과 일관성을 확인하기 위해 구성이 되어 있으며 주관식 문항의 경우 응답의 길이에 대해 파악하고자 구성되어 있다.

## 실험 방법

		설문 타입	
		온라인 설문조사	챗봇 설문조사
설문 주제	설문주제 #1 (스마트폰 과의존 실태 설문)	P1	P2
	설문주제 #2 (휴대폰 사용 전반에 관한 설문)	P2	P1

그림 8. 설문조사 타입 및 주제별 실험 assignment 방법

각 실험 참가자는 총 2회의 설문 요청을 받도록 되어있으며 2가지 type의 설문조사 방법에 대해서 각각의 설문주제는 중복되지 않도록 random하게 assign 되도록 하였다. 그리고 응답률의 확인을 위해서 2주일의 기간 동안 무작위 요일 및 시간에 설문조사 요청을 보내도록 하였다. 설문조사 완료 후 취합된 설문결과에 대해서는 Response Rate, Non-response item, Length of open-ended questions, Differentiation 을 측정하여 확인하였으며 설문조사가 완료된 후 모든 실험참가자와 반구조화 된 사후 인터뷰를 진행하도록 하였다.

## 실험 참가자

20대 ~ 40대 총 40명의 실험 참가자( $M_{age}=36.09$ ,  $SD_{age}=5.91$ ; 남성: 75%)는 IT 기업 내에서 Email을 통해 연구참여자를

섭외하였으며 연구참여자로 리쿠르팅 된 모든 인원은 Facebook messenger를 통해 ChatSurvey와 Personal Chat이 가능하도록 설정하였다. 참가자들의 챗봇에 대한 이해도나 사용경험에 대해서는 사후인터뷰를 통해 확인을 하였으며 5-point scale로 확인을 하였다.

- Chatbot에 대한 이해도(Awareness): Not at all - Slightly - Moderately - Very - Extremely (1 to 5)
- 사용 경험(Experience): Never - Rarely - Sometimes - Often - Always (1 to 5)

챗봇에 대한 이해도의 경우  $M_{\text{Awareness}} = 2.975$ ,  $SD_{\text{Awareness}} = 1.09$ 로 7명 정도의 실험참가자가 Chatbot에 대해 잘 모르는 것으로 나타났으나 대체적으로는 챗봇이 어떻게 동작하는지에 대해서는 이해하고 있는 것으로 파악되었다. 사용 경험의 경우  $M_{\text{Experience}} = 2.4$ ,  $SD_{\text{Experience}} = 1.1$ 로 일반적으로는 자주 사용하는 경우가 많지는 않으며 특별한 경우에 어쩔 수 없이 사용하게 된다고 답변하였다. 전혀 사용해보지 않았던 참가자도 25%인 10명이었고 본 실험을 통해 챗봇을 처음으로 경험해보았다고 답변하였다.

### 측정 방법

**Differentiation.** 이전 연구들에서도 응답 품질을 측정하기 위해 Differentiation의 값이 사용되어왔다. 각각의 설문조사에서 객관식의 경우 5-scale point로 실험이 진행되었으며 Differentiation에 대한 index를 계산하였고 [69] index가 클수록 실험 참가자가 선택한 답변의

differentiation이 있어 이에 대해 응답 품질이 더 높다고 판단할 수 있게 된다. [81] 이를 기준으로 웹과 챗봇 서버의 결과를 비교하여 연구 문제에 대해 검증을 해 보았다.

**Item Nonresponse.** 본 연구에서는 객관식/주관식 모두 입력을 하지 않고서는 Next Step으로 넘어가지 못하도록 디자인을 하였으며 객관식의 경우 모든 항목에 “응답안함”을 넣어 실험참가자가 “Don’ t Know” 답변을 할 수 있도록 하였고 이는 실험에 있어 No Response Rate에 대한 방법 별 차이의 Dominant한 결과를 확인하기 위하여 설계 되었다. [23] 객관식의 “응답안함” 항목을 포함하여 주관식에서의 의미 없는 답변 - “.”, “-“, “응답하지 않음” - 도 No Response 항목에 포함하여 그 결과를 비교하였다.

**Response Length of Open-ended Questions.** 설문 조사에서 응답 품질을 나타내는 또 다른 한가지 중요한 것은 주관식에 있어서 응답 길이이다. 이전 연구에서도 설문 방법의 응답 품질 비교에 사용된 척도이며 [53] 본 연구에서는 응답의 길이를 형태소를 기준으로 산정하였으며 실험 참가자가 사용하는 모든 형태소의 분석을 통해 응답에 대한 분석이 가능할 것이며 실험 참가자의 하나하나의 입력사항에 대해 모두 응답으로 간주하여 그 결과를 분석하기로 하였다

**Usability.** 유용성의 경우에는 실험참가자들이 실험을 모두 끝낸

뒤 진행한 사후 인터뷰를 통해 측정하였다. Usability를 측정하는 방법은 여러 가지가 있지만 [44] 그 중에서 Usefulness ( “It is useful” )에 대해 현재 대중적으로 가장 많이 사용되고 있는 웹 설문조사와 비교하여 챗봇 설문조사가 얼마나 유용한지에 대해 실험참가자에게 1 ~ 5 Point Scale로 물어보았고 이를 통해 챗봇을 통한 설문조사가 하나의 설문조사 방법으로 유용하게 사용될 수 있을지에 대해 가늠해보았다.

**Engagement.** 이전 연구에서도 확인 되었듯이 몰입도가 클수록 해당 Task에 대한 Enjoyment와 Experience에 기여를 한다고 볼 수 있다. [45] 설문조사에서도 참여자가 설문 Task 자체에 몰입할 수 있는 환경이 갖춰질 경우 응답 품질은 좋아지게 된다. 몰입도를 높이기 위해서는 여러 실험참가자가 사후 인터뷰에서 이야기 한 것처럼 참가자가 관심을 가질만한 설문조사 주제를 가지고 진행을 할 경우 가장 효과적이겠지만 본 연구에서는 각 실험참가자가 일반적으로 이해하고 있는 공통 주제를 가지고 설문을 진행하였고 단순히 설문 조사를 진행하는 인터페이스인 웹 혹은 챗봇의 방법을 비교하여 어느 방식이 더 몰입도를 가질 수 있는지에 대해 확인해 보았다.

**Response Time.** 여러 연구에서 이미 증명된 대로 Response Time이 짧을 경우에는 응답 품질이 나빠진다는 연구 결과가 존재 한다. [73, 11] 설문조사 문항에 대해 신중히 읽고 대답하는데 고민을 하면서 응답시간이 대체로 길어진다는 답변을 사후 인터뷰에서도 파악이

되었으며 주관식의 답변이 길어지면서 이로 인한 응답 시간도 더 소요가 되면서 응답품질은 더 올라가게 된다.

## 제 4 장 연구 결과

### 제 1 절 정량적 분석

#### 결과의 정규성 검증

결과의 정규성 가정을 충족하는지 판단하기 위해 Shapiro-Wilk의 정규성 검증을 실시하였다.

Shapiro-Wilk가 제안한 정규성 검증은 귀무가설이 정규성 가정을 만족하는 것이기 때문에, 유의확률이 .05 이상으로 나타나야 정규성 가정을 만족하는 것으로 본다. [51] 챗봇을 통해 진행한 객관적 사실의 주관식 유형에서 No Response는 분산이 없기 때문에 Shapiro-Wilk 통계량이 도출되지 않았고, 웹 설문 of 주관적 의견에 대한 주관식 유형의 Length 항목을 제외 하고는 모두 유의확률이 .05 미만으로 유의하게 나타났다. 유일하게 유의하지 않게 나타난 웹 설문의 주관적 의견에 대한 주관식 유형의 Length의 경우 챗봇에서는 유의한 결과를 보였기 때문에, 결과적으로 두 방법에서 모두 정규성 가정을 만족하는 항목은 없는 것으로 확인되었다.

즉 본 연구의 주요 변수들은 전반적으로 정규성 가정을 만족하지 못하는 것으로 판단되었기에, 통계적인 검증을 위해서는 정규성 가정을 필요로 하지 않는 비모수 통계를 활용하기로 하였다.



방법	내용	유형	항목	Shapiro-Wilk	df	p
챗봇	주관적 의견	주관식	Length	0.927*	40	.013
			No Response	0.147***	40	.000
		객관식	Differentiation	0.906**	40	.003
			No Response	0.243***	40	.000
	객관적 사실	주관식	Length	0.856***	40	.000
			No Response	—	—	—
		객관식	Differentiation	0.897**	40	.002
			No Response	0.393***	40	.000
	개인정보	혼합	No Response	0.365***	40	.000
웹	주관적 의견	주관식	Length	0.958	40	.138
			No Response	0.728***	40	.000
		객관식	Differentiation	0.935*	40	.023
			No Response	0.551***	40	.000
	객관적 사실	주관식	Length	0.902**	40	.002
			No Response	0.631***	40	.000
		객관식	Differentiation	0.880**	40	.001
			No Response	0.476***	40	.000
	개인정보	혼합	No Response	0.532***	40	.000

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

표 2. 결과의 정규성 검증

## 챗봇과 웹의 응답품질 비교

챗봇과 웹의 응답품질을 비교하기 위해 대응표본 비모수 통계 분석 방법인 윌콕슨의 부호순위 검정(Wilcoxon's signed rank test)을 실시하였다.

주관적 의견의 주관식에서는 Length( $Z=4.152$ ,  $p<.001$ )와 No Response ( $Z=4.585$ ,  $p<.001$ )가 모두 챗봇과 웹 간 유의한 차이를 보였다. Length의 평균을 비교해보면 챗봇은 19.84, 웹은 10.99로 챗봇이 높게 나타났고, 중위수도 챗봇은 16.90, 웹은 9.95로 챗봇이 더 높게 나타났다. No Response의 평균을 보면 챗봇은 0.10, 웹은 18.90으로 웹이 많이 높게 나타났고, 중위수를 봐도 챗봇은 0.00, 웹은 10.00으로 웹이 더 높게 나타났다. 즉 주관적 의견의 주관식 유형의 경우 챗봇이 웹보다 응답품질이 더 좋은 것으로 판단할 수 있다. 따라서 H1a 및 H1b는 성립됨을 알 수 있다.

주관적 의견의 객관식에서는 No Response가 챗봇과 웹 간에 유의한 차이를 보였다( $Z=2.801$ ,  $p<.01$ ). 평균 비교 결과 챗봇은 0.47, 웹은 3.67로 챗봇이 낮게 나타났고, 중위수는 모두 0.00이지만 상위 25% 수준을 파악하는 제3사분위수는 챗봇은 0.00이지만 웹은 3.13으로 나타나 웹 설문이 상대적으로 높았다. 즉 챗봇 보다는 웹이 주관적 의견의 객관식에서 무응답률이 상대적으로 높았다. H1c는 기각되었으며 H1d는 성립되었음을 알 수 있다.

객관적 사실의 주관식에서는 No Response가 챗봇과 웹 간에

유의한 차이를 보였다( $Z=3.787$ ,  $p<.001$ ). 평균 비교 결과 챗봇은 0.00, 웹은 12.50으로 챗봇이 낮게 나타났고, 중위수는 모두 0.00이지만 상위 25% 수준을 파악하는 제3사분 위수는 챗봇은 0.00이지만 웹은 25.00으로 나타나 웹 설문이 상대적으로 높았다. 즉 챗봇 보다는 웹이 객관적 사실의 주관식에서 무응답률이 상대적으로 높았다. 따라서 H2a는 기각되었으나 H2b는 성립됨을 알 수 있다.

객관적 사실의 객관식에서는 differentiation( $Z=2.581$ ,  $p<.05$ )과 No Response ( $Z=2.005$ ,  $p<.05$ )가 모두 챗봇과 웹 간 유의한 차이를 보였다. Differentiation의 평균을 비교해보면 챗봇은 68.01, 웹은 62.94로 챗봇이 높게 나타났고, 중위 수도 챗봇은 69.17, 웹은 65.60으로 챗봇이 더 높게 나타났다. No Response의 평균을 보면 챗봇은 0.33, 웹은 2.82로 웹이 많이 높게 나타났다. 즉 객관적 사실의 객관식 유형의 경우 챗봇이 웹보다 응답품질이 더 좋은 것으로 판단할 수 있다. 따라서 H2c와 H2d는 성립되었음을 알 수 있다.

개인 정보에 대한 No Reponse에는 유의차가 없으므로 H3은 기각되었다. 결과적으로 개인정보를 제외한 모든 유형에서 웹보다는 챗봇의 응답품질이 더 좋은 것으로 판단할 수 있다.

내용	유형	항목	집단	M±SD	Q1	Q2	Q3	Z	P
주관적 의견	주 관 식	Length	챗봇	19.84±10.08	12.25	16.90	25.96	4.152***	.000
			웹	10.99±7.28	5.49	9.95	15.50		
		Response	챗봇	0.10±0.63	0.00	0.00	0.00	4.585***	.000
			웹	18.90±25.85	0.00	10.00	23.00		
	객 관 식	Differ- entiate	챗봇	63.63±9.15	57.91	66.60	70.31	0.440	.660
			웹	62.15±10.02	57.13	65.23	69.58		
		Response	챗봇	0.47±2.07	0.00	0.00	0.00	2.801**	.005
			웹	3.67±7.52	0.00	0.00	3.13		
객관적 사실	주 관 식	Length	챗봇	4.80±2.14	3.25	4.06	5.50	0.840	.401
			웹	4.56±2.81	2.50	3.88	5.81		
		Response	챗봇	0.00M±0.00	0.00	0.00	0.00	3.787***	.000
			웹	12.50±20.80	0.00	0.00	25.00		
	객 관 식	Differ- entiate	챗봇	68.01±3.98	66.18	69.17	70.76	2.581*	.010
			웹	62.94±11.72	59.22	65.60	69.94		
		Response	챗봇	0.33±0.89	0.00	0.00	0.00	2.005*	.045
			웹	2.82±6.94	0.00	0.00	0.00		
개인 정보	혼 합	No	챗봇	3.89±12.21	0.00	0.00	0.00	0.821	.412
		Response	웹	7.28±16.08	0.00	0.00	10.00		

\*  $p<.05$ , \*\*  $p<.01$ , \*\*\*  $p<.001$

표 3. 챗봇과 웹의 응답 품질 차이 검증 결과

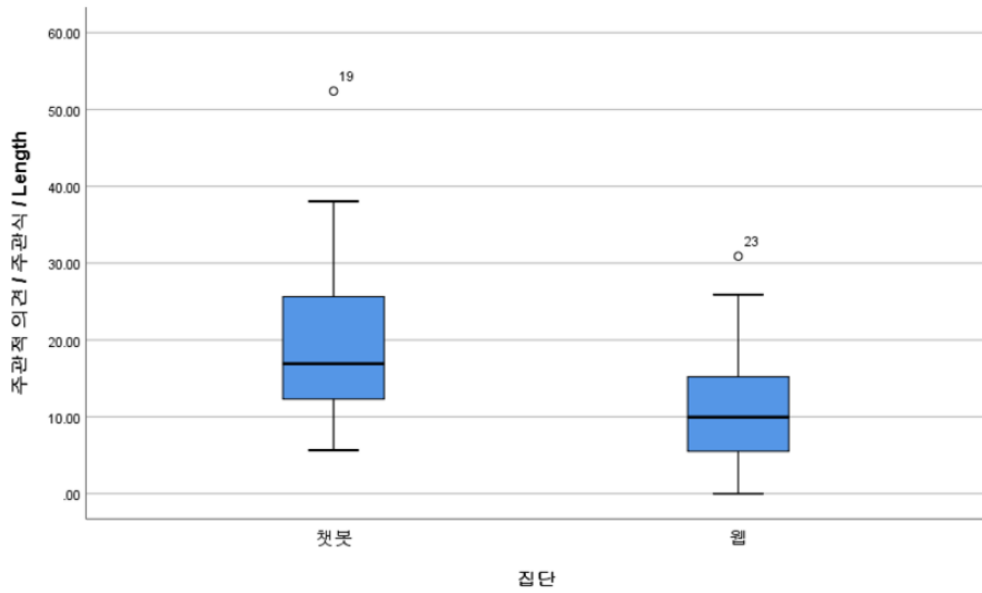


그림 9. 챗봇과 웹의 주관적 의견 / 주관식의 Length

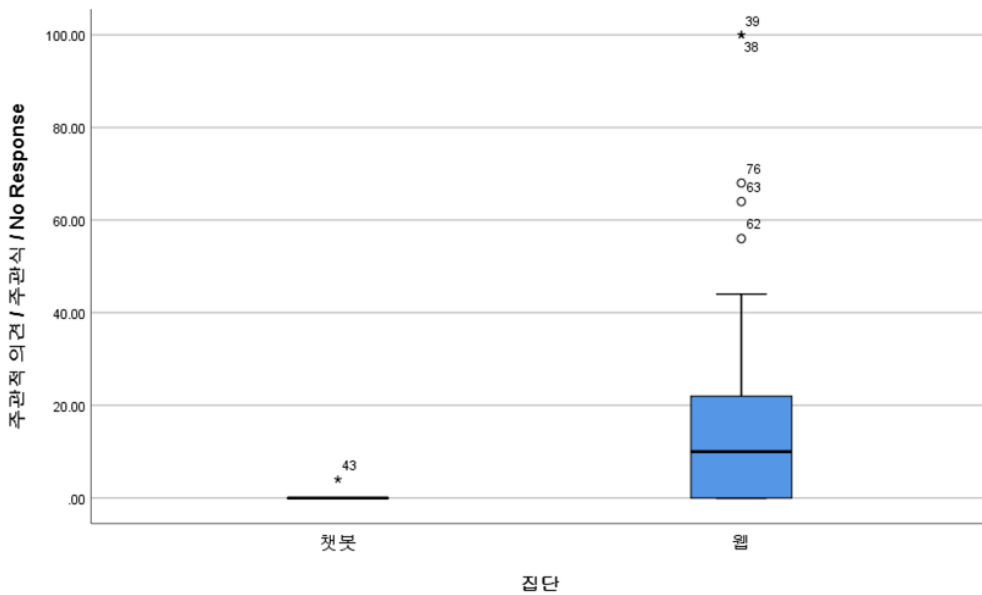


그림 10. 챗봇과 웹의 주관적 의견 / 주관식의 No Response

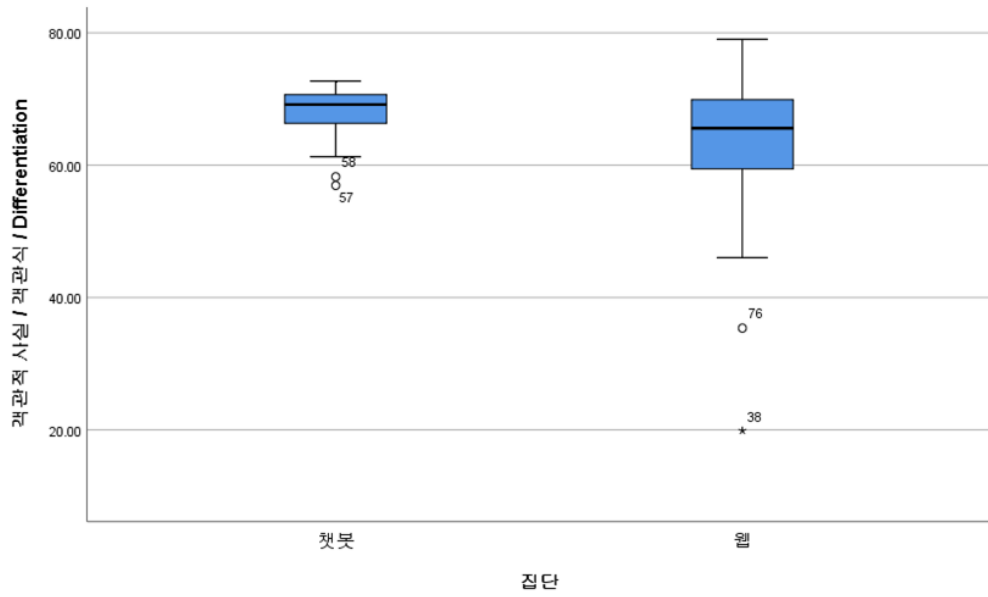


그림 11. 챗봇과 웹의 객관적 사실 / 객관식의 Differentiation

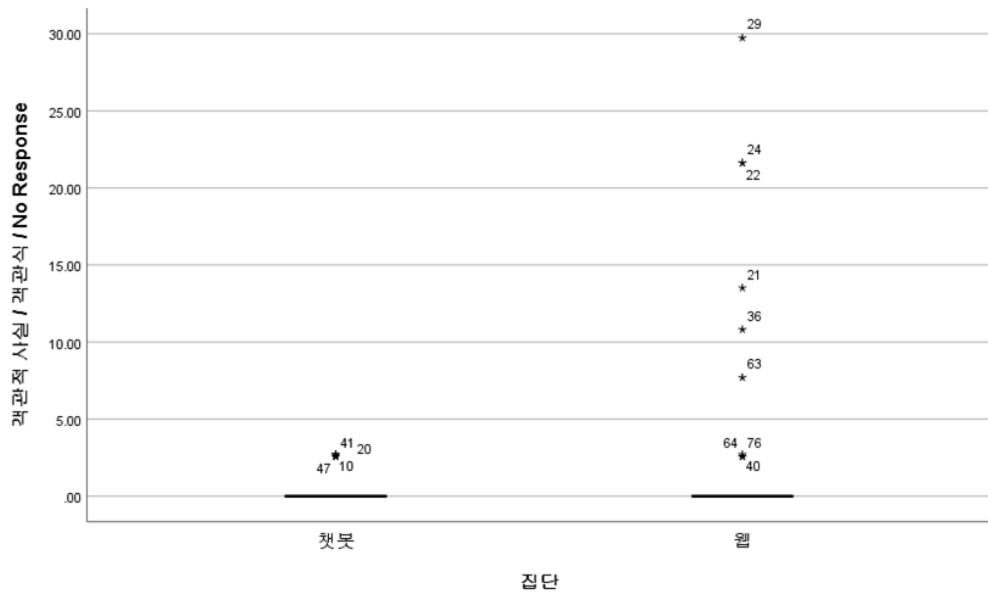


그림 12. 챗봇과 웹의 객관적 사실 / 객관식의 No Response

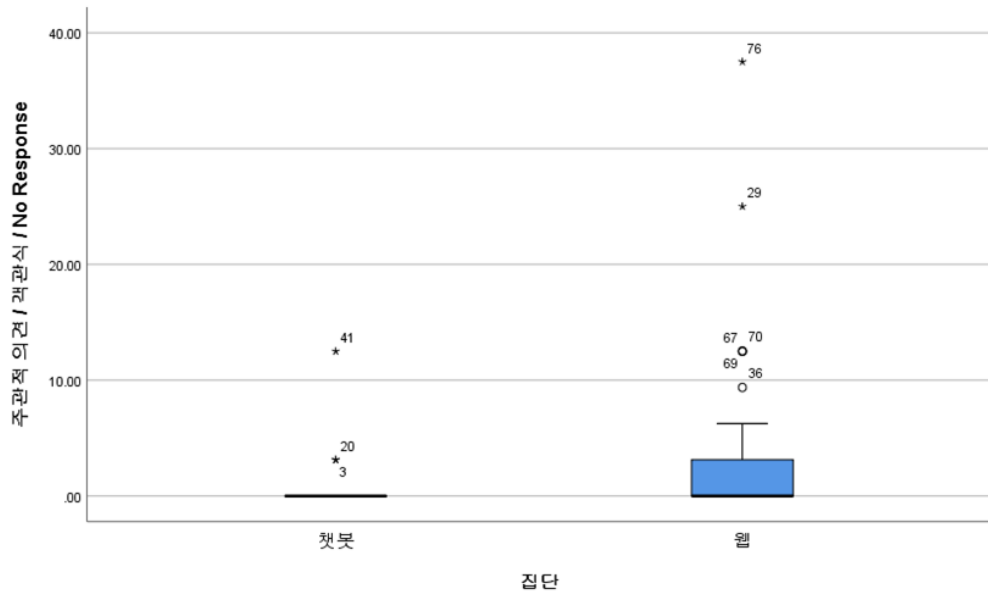


그림 13. 챗봇과 웹의 주관적 의견 / 객관식의 No Response

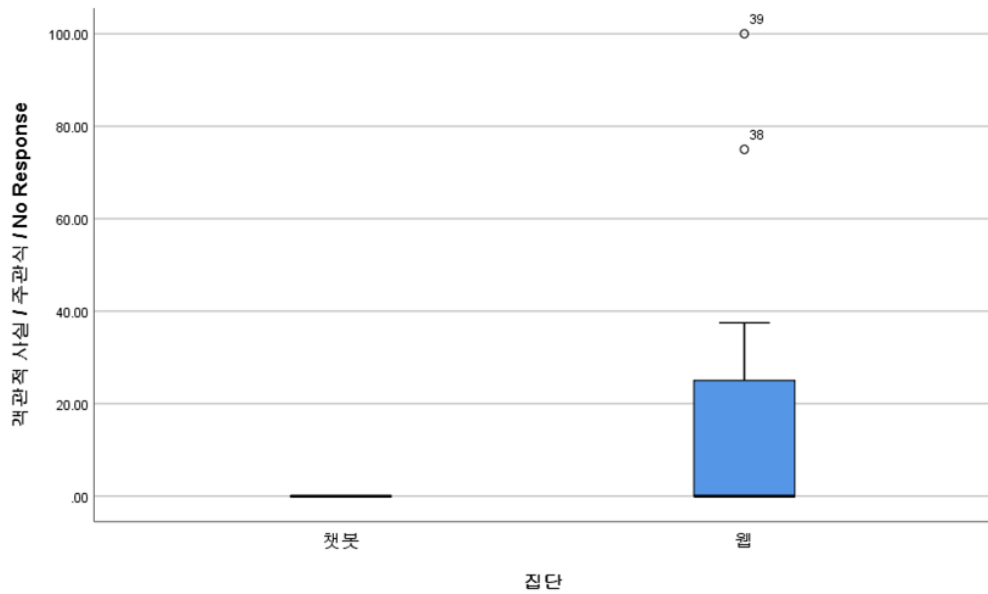


그림 14. 챗봇과 웹의 객관적 사실 / 주관식의 No Response

## 유용성과 몰입감

실험참가자들은 챗봇과 웹 설문조사를 모두 경험을 하였고 이러한 경험을 바탕으로 사후 설문 시에 챗봇 설문조사와 웹 설문조사의 유용성 및 몰입감에 대해서 5 point scale로 답변을 받아 분석하였다.

챗봇과 웹의 유용성과 몰입감의 차이를 검증하기 위해 윌콕슨의 부호순위 검정 (Wilcoxon' s signed rank test)을 실시하였다. 그 결과 유용성 ( $Z = 3.678$ ,  $P < .001$ )과 몰입감 ( $Z = 4.452$ ,  $P < .001$ ) 모두 챗봇과 웹이 유의한 차이를 보였다.

항목	집단	M±SD	Q1	Q2	Q3	Z	P
유용성	챗봇	3.88±0.72	3.00	4.00	4.00	3.678**	.000
	웹	3.20±0.88	3.00	3.00	4.00		
몰입감	챗봇	3.78±0.83	3.00	4.00	4.00	4.452***	.000
	웹	2.58±0.96	2.00	3.00	3.00		

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

표 4. 챗봇과 웹의 유용성과 몰입감 비교

평균 비교 결과 유용성은 챗봇 ( $M = 3.88$ )이 웹 ( $M = 3.20$ )보다 높게 나타났고, 중위수도 챗봇 (4.00)이 웹 (3.00)보다 높았다. 몰입감 평균도 챗봇 ( $M=3.78$ )이 웹 ( $M=2.58$ )보다 높았고, 중위수도 챗봇 (4.00)이 웹 (3.00)보다 높았다. 결과적으로 유용성과 몰입감 모두 웹보다는 챗봇이 더 좋은 것으로 판단할 수 있다. 따라서 H4와



H5의 가설은 성립되었음을 알 수 있다.

## 거부감

모든 실험 참가자들이 챗봇을 이용한 설문조사는 처음이었으며 친근한 말투의 사용으로 인해 설문조사 자체가 구어체로 이루어져 거부감이 느껴질 수도 있을 것이다. 그래서 메신저라는 인터페이스를 통해 챗봇으로 진행이 되는 설문조사에 대한 거부감을 사후 인터뷰에서 5-Point Scale로 실험 참가자와 확인해 보았다.

Q: 챗봇으로 설문조사를 진행해보고 새로운 설문 방법인 챗봇 설문조사에 대한 거부감이 있었는지 그 정도를 1~5점 (1점: 전혀 거부감이 없음, 5점: 심각하게 거부감이 듬)으로 알려주시고 그 이유를 말씀해주세요.

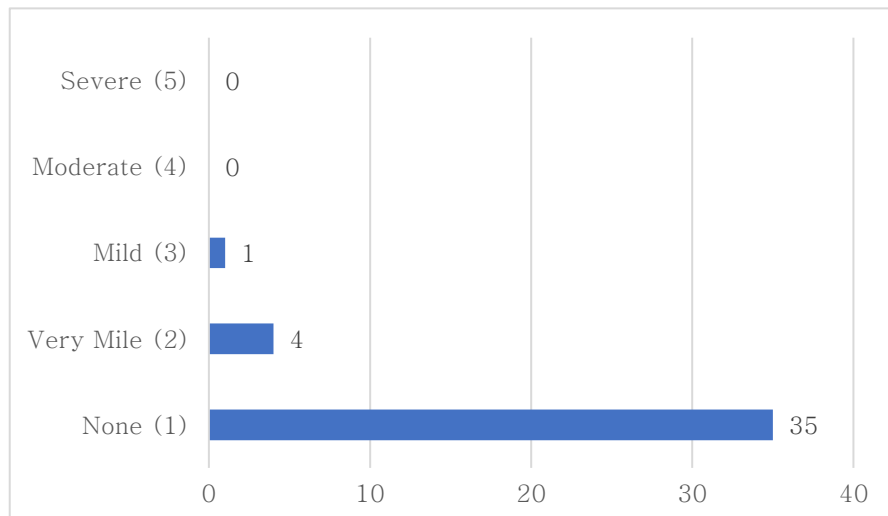


그림 15. 챗봇 설문조사에 대한 거부감

거부감에 대한 결과는 거부감<sub>avg</sub> = 1.15, 거부감<sub>sd</sub>=0.43으로 아주 낮게 나타났으며 대부분의 실험 참가자는 전혀 거부감이 없다고 대답하였다. 40명 중 35명은 전혀 거부감이 없다고 대답을 하였고 그 이유는 이미 친숙한 메신저라는 도구를 이용해 대화하는 것에 불과한 것이라 거부감이 없고 구어체의 질문에 대해서도 오히려 친구 같은 편안함을 느낄 수 있다고 답변을 하였다. 약간 거부감이 느꼈다고 답변을 한 실험 참가자는 4명이 있었으며 새로운 환경에 대한 거부감이나 혹시 모를 개인정보 노출에 대한 거부감일 뿐 심각한 거부감은 아니라고 의견을 전달하였다. 거부감을 느꼈다는 실험 참가자는 1명이 있었으나 이는 챗봇 설문조사에 대한 거부감이라기 보다는 설문에 대한 기본적인 거부감을 뜻한다고 하여 실제 챗봇을 통한 설문 조사의 경우 사용자로부터 거부감으로 인해 부작용은 없다고 봐야 할 것이다.

## 제 2 절 정성적 분석

실험 참가자와 진행 한 사후 설문지 답변에 대해서는 Thematic Map을 통해 분석을 진행하였으며 아래와 같이 크게 5가지로 분류가 가능하였다.

### 새로운 설문조사 방법으로 인한 신선함/새로움

실험 참가자들의 대다수가 챗봇으로 설문으로 진행하는 것에 대해 ‘신선하다’ 혹은 ‘새롭다’ 라는 반응을 보여주었고 이는 챗봇이 새롭다라는 것보다는 챗봇을 통해 상호 작용을 하면서 설문조사에 이용하는 것에 대한 신선함, 새로움을 표현한다고 하였다. 우리가 잘 알고 있는 사용자 인터페이스라도 그것을 현재 사용하고 있지 않은 용도로 사용할 경우 익숙함을 넘어서는 새로운 경험을 사용자에게 안겨줄 수 있다는 것을 의미한다. 새로운 경험이 사용자로 하여금 해야 할 숙제가 아닌 즐거움을 줄 수 있다는 것을 확인하였으며 이를 통해 설문조사에 더욱 적극적으로 임할 수 있게 된다고 하였다.

(P35) “챗봇으로는 설문조사를 해본 것이 처음이기 때문에 새로운 면이 있었다. 챗봇이 다양한 방면으로 사용되는 것 같은 느낌을 받았으며 사용자에게 관심을 가질 수 있는 기술이 설문조사에 적용됨으로써 재미있게 설문을 진행할 수 있었다”. (P33) “온라인

설문조사는 시험지에 답을 쓰는 것 같았지만, 챗봇으로 설문조사를 진행하다 보니 신선함이 많이 느껴졌고 실제 인터뷰 하는 느낌을 받게 되었고 그래서 더욱 성실하게 답변을 했던 것 같다. 그래서 시간이 오래 걸린다고 생각하지도 않았던 것 같다.”

결국 챗봇도 설문조사의 한 가지 중요한 방법으로 효과를 발휘할 수 있겠지만 사용자에게 새로움, 즐거움을 지속적으로 전달 할 수 있는 매개체가 되어야 하며 이에 대한 추가적인 연구는 필요하다.

### 익숙한 메신저 인터페이스를 통한 편안함/편리함

모바일 인덱스<sup>14</sup>에 따르면 2020년 5월 현재 카카오톡 월 사용시간이 3개월 연속 4억시간을 넘었다고 한다. 그리고 카카오톡에 따르면 2020년 1분기 평균 월간활성이용자수(MAU)는 4,519만명을 기록했고 이는 스마트폰을 사용하는 모든 국민들이 사용하는 국민 메신저라고 할 수 있다. 카카오톡뿐 만 아니라 라인, 페이스북 메신저 등 메신저라는 플랫폼은 어린 학생부터 고령층까지 메신저라는 형태에는 익숙해져 있는 것이 현실이다. 소비자로서의 가치 구조를 규명한 연구에서도 확인되었듯이 “대중성” - “사용의 익숙함/편리함” - “편안함/안정감”의 구조를 가지게 된다. [6] 그러다 보니 메신저라는 인터페이스를 통해 챗봇의 기술이 결합되어 설문조사를 진행하면서 설문 참가자들이 느끼는 가장 중요한 요소도 ‘익숙함’, ‘편안함’이라는

---

<sup>14</sup> 모바일인덱스HD (<https://hd.mobileindex.com/>)

답변이었다. 설문 조사에서도 설문하는 방법이 불편 할 경우 그 효과가 반감 됨에 비추어 설문 참가자들에게 익숙한 인터페이스를 통해 설문 조사가 진행되어 그 답변을 진행함에 있어서 훨씬 여유롭게 깊게 생각이 가능하다는 의견이 많았다.

*(P39) 기존에 사용하던 익숙한 interface를 통해 진행하다 보니 흥미롭게 진행할 수 있었습니다. (P33) 통화나 다른 서베이에 비해 부담이 적어서 좋았고, 생각날 때 언제든지 할 수 있는 장점이 있는 것 같습니다. (P11) 요즘 세대에 적합한 방식으로 설문이 진행되고 언제 어디서나, 그리고 참여하는 사람이 시간이 날 때 진행할 수 있는 장점이 있을 것 같습니다.*

이전에 다양한 연구에서도 확인 되었지만 설문 조사에 대한 응답률은 설문 방법에 관계없이 고령자가 가장 높게 나온다는 것이다. [78] 2017년 와이즈앱<sup>15</sup>에서 조사한 바에 따르면 세대별 가장 오래 사용하는 앱을 조사했을 때 10대/20대는 유튜브이지만 50대 이상에서는 카카오톡이 84백만 시간으로 가장 많은 사용시간을 보였다. 이렇게 메신저를 통한 설문 조사의 경우 face-to-face 혹은 전화 인터뷰 보다는 비대면을 선호하는 젊은 세대의 취향에 적합할 뿐만 아니라 50대 이상의 Senior 그룹에서도 가장 오래 동안 사용하고 익숙한 메신저를 사용하게 된다면 전 연령에서 한가지 설문 조사 방법을 통해 설문을 진행할 수 있다고 보여지고 새로운 기술에 의한 고령자의 이탈현상은 적으리라 본다. 본 연구에서는 50대 이상의 설문 참가자가

---

<sup>15</sup> 와이즈앱 (<https://www.wiseapp.co.kr/>)

없어 그 현상을 분석할 수는 없었지만 이후 연구에서 다뤄야 할 주제로 보여진다.

(P36) 실험을 하면서 생각이 들었지만, 메신저를 통해 설문 조사를 하는 것은 요즘 카톡을 많이 사용하시는 부모님들에게도 쉽게 통할 수 있는 방법이라고 생각했습니다.

메신저를 PC에서 사용하는 사람도 있기는 하지만 이동 중에는 모바일 디바이스를 통해 사용하게 되고 전통적인 설문 조사 방법의 경우 대면일 경우 상대방이 기다리는 상황으로 인해 조급함으로 인해 응답의 품질은 낮아질 수 있으며 온라인 설문조사의 경우 중간에 Session이 끊어져 기존 진행했던 설문 응답을 모두 잃을 염려도 있고 다른 링크에 의해 진행했던 내용들을 잃어 버릴 수 있다는 불안감이 큰 것이 반해 메신저는 실험 참가자들이 언제나 원할 때 원하는 장소에서 설문을 이어나갈 수 있다는 것에 대해 크게 반기는 것을 알 수가 있었다. 중단된 설문을 이어나가기 위한 적절한 알림만 주어진다면 설문에 대한 응답률을 높일 수가 있다.

(P20) 요즘 세대에 적합한 방식으로 설문이 진행되고 언제 어디서나, 그리고 참여하는 사람이 시간이 날 때 진행할 수 있는 장점이 있을 것 같습니다. (P31) 메신저를 통해 모바일로 보통 진행할 것 같아서 접근성, 사용성이 좋은 것 같다. (P22) 설문 조사를 하면서 여러 번 그 결과를 날려본 입장에서는 메신저로 하게 된다면 그 내용이 절대 사라지지 않을 것이므로 같은 설문 조사를 2번 이상 답하는 경우가 없을 것 같아 마음에 듭니다. (P24) 대화형 서비스처럼 보이는데, 특히 설문이 긴 경우 시간적 여유를 두고 진행할 수 있는 점이 편했습니다.

## 상호작용이 있는 대화형 설문에 의한 높은 몰입감

Conversation Agent의 경우 여러 가지의 인터페이스가 존재한다. Siri, Alexa, Google Assistance와 같이 음성 대화 인터페이스를 가진 형태가 있고 챗봇과 같이 Text Base의 인터페이스를 가진 형태도 있고 실제로는 발화하여 대화하지는 않지만 text를 기반으로 메신저를 통해 서로의 의견을 전달하는 챗봇 또한 대화형이라는 범주에 속하게 되고 이러한 특징에 대해 가장 인상적이라는 답변을 대다수의 실험참가자들이 언급하였다. 이전 연구에서 확인한 바대로 상호 작용의 수준이 높을 경우 사용자 경험에 긍정적인 효과를 가져다 주게 된다고 한다. [3] 실제로도 사후 인터뷰에서 확인한 몰입도의 경우에도 챗봇의 몰입도가 더욱 높은 이유는 대화형 인터페이스를 사용하였기 때문이라는 의견이 지배적이었다. 대화형으로 설문이 진행되었기 때문에 지루함이 없었고 실제 인터뷰를 하는 느낌 혹은 친구와 대화를 하게 되는 느낌을 받게 되어 답변에 대해서도 신중하게 진행하게 되고 열심히 하게 된다는 의견들도 많았다.

(P1) 챗봇을 통한 설문 조사는 대화형으로 자연스럽게 대화하듯이 설문을 하여 설문자가 마음을 조금 더 편하게 가지고 나아가 더 솔직하게 대답할 수 있게 하는 것 같다. (P6) 질문에서의 차이는 없었으나 대화를 하다 보니 설문에 대한 집중을 좀 더 하게 되었다. (P12) 챗봇의 경우 대화하는 느낌이라 성의 없게 대답하기는 어려울 듯하다. (P14) 전체적인 질문 자체에 대해서 상대방과 대화하는

방식이다 보니 성실히 답하려고 한 것 같습니다. (P20) Chatbot 이 실제 대화를 거는 것 같아 질문에 대한 대답이 자연스레 떠올라 주저 없이 답변한 것 같다.

실험에서는 챗봇의 경우 대화형으로 설문이 진행되기는 했지만 온라인 설문 조사와 동일한 질문 항목이 나열되어 있었다. 이 외에 대화형의 느낌을 주기 위해 실험 참가자의 의견을 입력한 후에는 그에 대한 반응 - “아, 그랬구나” , “나도 그런 것 같아” - 라는 어구를 일부 추가한 차이점이 있었다. 이러한 설문의 방법으로 인해 실험 참가자들은 챗봇이 Interactive 하게 느껴지거나 설문 답변에 따라 설문 문항이 달라지면서 feedback이 있는 것처럼 느껴졌다고 하였다. 그리고, 한 번에 많은 설문 문항이 보이는 온라인 설문 조사에 비해서 챗봇을 통해 설문을 진행하게 되니 실시간으로 설문을 진행한다는 느낌을 받게 되었다고 하였다. 위에서 언급한 효과에 의해 대화형이라는 인터페이스는 사용자와의 상호작용을 통해 몰입감을 증가 [19] 시키는 원인이 되며 설문을 더욱 즐겁게 할 수 있게 되어 결과적으로는 응답 품질이 좋아지게 되는 가장 큰 요인이 된다고 할 수 있다.

(P22) 챗봇 설문 조사는 순차적으로 질문하는 것이 아니라, 이전 제 대답에 따라서 질문이 바뀌는 것으로 느껴지는 것이 차이가 있었습니다.

(P34) Chatbot은 대화하는 방식이므로 feedback을 바로 받을 수 있는 장점으로 몰입도가 더 큰 것으로 생각됩니다. (P16) 챗봇으로 설문 조사를 했을 때에는 실제로 대화하듯이 설문조사를 한 것 같다. (P21) 온라인 설문 조사는 정해진 질문을 하고 있다면 Chatbot은 상황과



흐름에 맞게 질문하는 것 같이 느껴졌다.

### 친구와 같은 친근한 말투에 의한 능동적이고 솔직한 반응

이전 연구에서도 확인 되었듯이 온라인 설문 조사의 경우 친근한 말투로 인한 응답 품질의 차이는 없지만 챗봇에서는 친근한 말투에 의해 그 응답 품질이 좋아짐을 확인 할 수 있었다. [2] 결국 챗봇을 사용함에 있어 가장 큰 이점은 구어체를 사용하더라도 거부감이 발생하지 않는 것에 있고 이로 인해 응답 품질에 큰 영향을 주게 되는 것이다.

친근한 말투로 인해 실험 참가자들은 딱딱한 기존 설문 조사 방식에서 벗어나 설문 조사에서 친절함을 느끼게 되었고 친구와 같은 편안함을 느끼기 되었고 특히 실제로 인터뷰를 하는 듯한 느낌도 드는 경우가 있다고 하였다. 이런 감정들은 설문 답변을 더 솔직하고 능동적으로 할 수 있게 만드는 것 같다고 하였다.

(P11) 챗봇이 대화형이다 보니 친구와 대화하듯이 하게 되어 한번 더 생각하게 되는 것 같습니다. (P24) 친근한 말투로 대화하는 듯한 느낌이 들었습니다. (P30) 일반 설문보다는 사람과 설문하는 느낌이 들고 친절하게 대화하는 느낌이 들어 설문을 더 잘하게 되는 것 같습니다. (P37) 친근한 말투의 대화형식이라 질문 받는 느낌이 좀 더 인간미가 느껴졌습니다. (P35) 전체적인 설문에 있어 챗봇이 다소 사람 같이 느껴져 일반적인 딱딱한 설문보다는 친근한 감정이 생기게 되고 질문을 더 꼼꼼히 읽게 되고 답변하게 되는 것 같습니다. (P5) 어느 부분에서는

사람과 대화하는 느낌도 있었으며 챗봇인 것을 인지하면서도 대화를 재미있게 하고 있었습니다. (P15) 답변에 대해 다음 단계로 진행될 때, 내가 사람과 커뮤니케이션을 하고 있는지 깜짝 깜짝 놀랄 때가 있었습니다. (P32) 구어체로 진행되던 점이 아주 좋았던 것 같습니다. 딱딱하지 않고 재미 있었으며 답변에 대해 오류가 있을 때 기다려 달라고 하는 말투 등도 좋았습니다. (P26) 약간 어색한 부분도 있었지만 만약 말투를 좀 더 다듬는다면 챗봇인지 상대방이 사람인지 인지하기 어려울 수도 있어 보입니다.

### 높은 모바일 접근성에 의한 낮아진 시간/장소 제약

본 연구의 실험 참가자들은 온라인 설문조사와 챗봇 설문 조사 모두 모바일 디바이스에서 진행하였다. 챗봇의 경우 메시지의 형태이기 때문에 모바일 디바이스에서 대부분 진행이 되는 점이 있긴 하지만 온라인 설문조사의 요청도 메시저로 링크가 되도록 설계를 하였다. 두 실험의 설문 도구의 차이를 없애기 위한 방법이었고 이로 인해 모든 실험 참가자가 동일한 모바일 환경에서 테스트를 진행하게 되었다.

온라인 설문조사와 챗봇 설문조사는 모두 모바일 디바이스 혹은 PC에서도 가능함에도 불구하고 실험 참가자들은 챗봇 설문조사는 모바일 디바이스, 온라인 설문조사는 PC에서 진행하는 것이라는 생각을 가지고 있었고 각각의 방식이 더 적합하다는 의견에 대체로 동의하였다. 모바일을 사용하게 됨으로써 설문의 시간/공간적인 제약이 낮아지게 되었고 실험 참가자에게 설문 장소를 확인 한 결과 설문을 진행하는

장소와 시간은 아주 다양한 것으로 확인 되었다. 진행한 설문 조사의 28.8%가 고정된 장소 (예, 집, 회사, 학교 혹은 카페)가 아닌 이동 중에 설문조사를 진행한 것으로 파악되었다.

(P16) 챗봇으로 서버이를 진행하는 것은 기존 웹 서버이 대비 모바일에서 사용하기 훨씬 편했다. (P21) 챗봇을 통한 설문 조사는 모바일로도 쉽게 할 수 있고, 기존의 설문조사 보다 덜 지루하다. (P28) 사용의 확장성 측면에서 챗봇 서버이는 모바일에서 쉽게 사용할 수 있어 좋은 것 같다. (P31) 챗봇 설 조사는 메신저를 통해 모바일로 보통 진행할 것 같아서 접근성, 사용성이 좋은 것 같다

그리고 모바일을 사용하게 되는 환경적인 이유로 인해 모바일 키보드나 화면 사이즈의 제약으로 인한 불편함에 대해서도 설문 참가자에게 확인해보았지만 모바일에 대한 불편함을 호소하는 인원은 모두 13% 정도로 낮게 나타났다. 물론, 실험참가자가 모바일에 익숙한 세대인 20 ~ 40대에서 설문이 진행되었기 때문에 모든 연령대를 대변할 수는 없는 것은 사실이다.

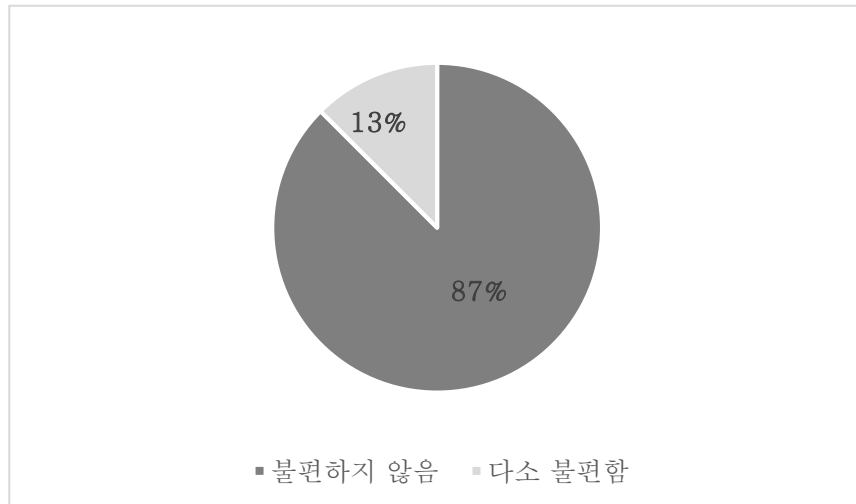


그림 16. 설문 시 모바일 키보드 사용의 불편함 조사

다소 불편하다고 응답한 5명의 실험 참가자(P2, P6, P7, P15, P40)들은 긴 답변을 진행하게 될 때 오타 등의 이유로 다소 불편한 점은 있다는 의견을 주었습니다.

(P22) 모바일로 설문조사를 진행함에 있어 길게 답변해야 할 경우 다소 귀찮기는 하지만 그렇다고 답변을 일부러 짧게 하게 되지는 않았습니다.

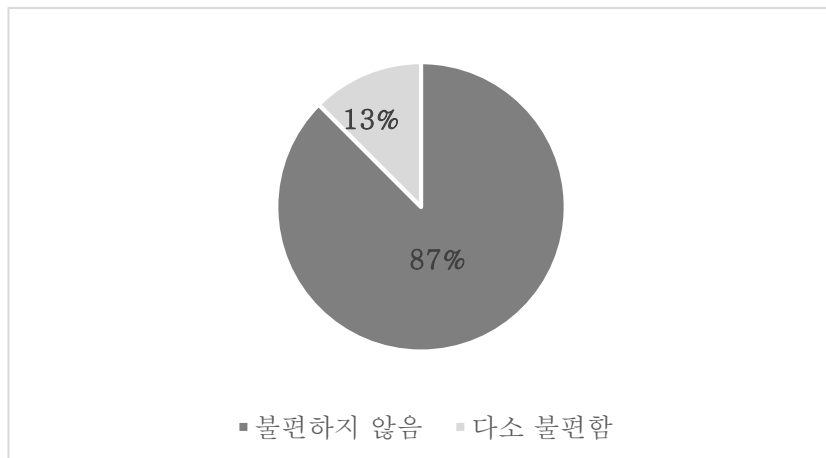


그림 17. 설문시 화면 사이즈 제약으로 인한 스크롤의 불편함

설문 시 스크롤로 인한 불편함에 대해서는 5명(P5, P13, P16, P23, P30)의 참가자가 다소 불편하다는 의견을 전달하였고 이는 챗봇 설문조사 디자인 시 꼭 고려해야 할 사항으로 생각된다.

*(P5) 모바일에서 챗봇을 통해 설문을 진행하면서 스크롤을 가끔 하게 되는데 그 부분은 불편했었습니다. 전반적인 것을 다 보고 싶는데 내려서 보니 빠지는 부분이나 읽다가 몰입을 방해할 때도 있었던 것으로 기억합니다.*

### 제 3 절 챗봇 설문 조사 디자인을 위한 고려 사항

챗봇으로 설문 조사를 하게 될 경우 모바일 디바이스를 통한 설문이 이루어지게 될 것이므로 챗봇이라는 기술 이용에 대한 고려사항과 모바일 사용에 의한 불편한 점 등을 개선하는 디자인이 필요하다.

#### 높은 상호 작용을 위한 설문 디자인

챗봇의 가장 기본적인 특징이자 실험 참가자들이 가장 인상적으로 생각하는 것이 바로 대화형으로 interaction을 통해 피드백을 받을 수 있다는 것이다. Interaction이라는게 본 실험에서는 “그렇구나”, “나도 비슷한 생각이야”, “아니라고 생각하는구나” 등의 간단한 리액션을 포함하였지만 실험 참가자들이 생각하는 것은 그것 보다는 한 단계 더 진화된 것을 원하는 것을 알 수 있었다. 실제로 실험 참가자들이 설문 항목에 대해 물어보거나 이해가 되지 않는 부분에 대해서는 추가적인 요청을 하고자 할 경우가 있었으나 그런 것에 대한 대응은 불가능 하였다. 이는 자연어 처리나 머신 러닝 그리고 챗봇 기술의 발전이 수반되어야 하겠지만 챗봇 디자인할 때에는 이 부분에 대해서는 고려하여 현재 가진 기술을 기반으로 더욱 자연스러운 대화가 가능하도록 되어야 하며 사용자의 질문을 처리할 수 있는 방안을 마련해 두는 것이 필요하다.

(P13) 좀 더 인간의 말투와 유사하게 만들고 Q&A를 다양하게 수용하게 하면 좋을 것 같습니다.

그리고 본 실험에서는 온라인 설문 조사와의 차이점을 파악하기 위해 설문 조사의 형식을 그대로 챗봇으로 옮긴 것이 불과하였다. 그렇게 디자인이 되었기 때문에 대화형의 장점을 모두 활용하지는 못하였다. 실험 참가자의 답변에 기반한 Interaction이 중요하고 이를 기반으로 스토리라인을 구성하여 설문을 진행하거나 설문의 형식을 띄는 게 아니라 대화하면서 자연스럽게 설문에 대한 답변을 받을 수 있도록 구성할 수 있다면 응답 품질은 더욱 올라갈 것이다.

(P9) 개인적인 생각이지만 획일적인 질문이 아니라 이전 답변에 기반해 너에게만 이 질문하는 것이라는 것을 제대로 인지할 수 있게 하면 좋을 것 같아요. (P4) 일반적인 Survey 보다 대화형식으로 이루어지다 보니 지루함이 덜 했으며, 스토리가 이어지는 느낌을 받아 설문에 적극적으로 임해지는 느낌이 들었습니다. (P12) 설문을 직접적으로 하는 게 아니라 대화를 하면서 설문의 답변을 받을 수 있도록 만들면 좋을 것 같습니다.

실험 참가자들은 피드백에 대한 언급이 많았다. 설문 조사의 경우 설문 조사만 할 뿐이고 피드백을 받아본 적이 없었고 그로 인해 일반적으로 설문 조사에는 응하지 않는다고 하였다. 결국 본인의 진행한 설문 조사의 전체 결과나 이로 인해 개선된 점들에 대한 피드백이 챗봇을 통해서도 더 잘 표현 될 수 있다라고 생각하였다. 결국 본인이 투입하는 설문 조사를 진행하는 시간이 의미가 있고 본인에게 도움이 된다고 느껴질 때 그 설문 조사에 대한 응답률과 응답품질은 높아지게 되고 그 소통을 할 수 있는 창구가 챗봇이 되는 것이 가장 적당하다고

보여진다.

(P21) 진행한 Survey 관련 정보를 향후 추가 정보를 add하여 안내문 형식으로 채팅 창을 통해 자동 수신 받을 수 있다면 좋을 것 같습니다. 가령, Autonomous 관련 설문 시 답변이 Autonomous 의 Mobility 산업 관련 이었다면 향후 관련 정보를 안내 받는 형식이면 응답률이 높아지지 않을까요? (P24) 한국사람은 소속감을 중요시하는데, 본인의 답변이 전체 참석자중 몇%에 속하는지 바로 바로 나오면 재미있을 것 같아요.

## 의인화와 개인화

실험 참가자들이 챗봇으로 설문 조사를 진행할 때 마치 대화하는 것 같다는 의미가 실제 사람과 대화를 한 것 같다는 의미는 아니었다. 물론 사람과 대화하는 것과 같이 느껴졌다는 응답도 있었으나 많은 실험 참가자들은 기계적이지 않고 더욱 더 자연스러운 대화를 원하는 것으로 나타났다. 대화 상대가 컴퓨터가 아니라 사람으로 생각할 경우 참가자들이 더 많은 말을 하고 시간을 쓴다는 연구 결과[70]에서 확인되었고 챗봇의 특성으로 의인화와 개인화의 수준이 높을 때 서비스에 대한 수용의도가 높아진다는 점도 확인되었다. [4] 챗봇에서 의인화와 개인화의 수준을 높이게 된다면 설문 조사에 대한 응답 품질이 좋아지게 될 것이다. 결국 챗봇 설문 조사에서 적용 가능한 의인화, 개인화의 요소를 발굴하여 응용하는 것이 중요하다. 스마트 홈 인터페이스의 의인화의 효과에 대한 연구에서 [5] 의인화를 위한



5가지의 요인을 참고하여 [57] 실험을 진행하였으며 이로 인해 친밀감, 호감도가 상승했다는 결과를 확인하였다. 이와 마찬가지로 챗봇의 의인화를 위해 이 5가지를 참고하여 대화를 구성하는 것이 필요하다. 그 다섯 가지 요인은 감정적 대응 (Emotional Responsibility), 따뜻함 (interpersonal warmth), 인지적 개방성 (cognitive openness), 독립적 능동성 (individual agency), 깊이 (depth)이다. 각각의 연구에서 사용한 실험의 예시는 아래와 같으며 이는 설문 조사를 설계 시에 참고하면 된다.

요인	적용 속성	HN High (보니)	HN low (신디)
감정적 대응	감정 어휘	"네~! 따뜻한 무드등으로 켜 드릴게요^^"	"무드등 모드로 조절합니다."
따뜻함	부드러운 어투	"20도로 설정되어 있어요."	"20도로 설정되어 있습니다."
인지적 개방	질문	"지금 방 안의 온도는 어떠신가요?"	없음
독립적 능동성	제안, 추천	"오늘은 날씨가 좋네요~ 오늘 같은 날씨에 어울리는 음악을 들어볼까요?"	없음
깊이	추가적 피드백	"네~ 공기 청정 기능을 작동시킬게요. 금방 쾌적해질 거예요!"	"공기청정 기능을 가동합니다."

그림 18. NH수준의 메시지 조작 예시 [5]



외형	의인화 고			의인화 저		
						
이름	안나	제니	리아	쇼핑봇A	쇼핑봇B	쇼핑봇C
상호작용 방식	인간적 본성의 구성요인		적용 속성	의인화 고		의인화 저
	감정적 대응		감정 어휘, 이모티콘	무엇을 도와드릴까요?^^		무엇을 도와드릴까요?
	따뜻함		부드러운 어투	고객님의 '성별'을 선택해주세요.		고객님의 '성별'을 선택해주세요.
	독립적 능동성		제안, 추천	OO와 함께 쇼핑하실래요?		OO와의 쇼핑을 시작합니다.
	깊이		추가적 피드백	요즘 인기 있는 '20대 여성 셔츠/블라우스'를 보여드릴게요. 마음에 드셨으면 좋겠어요^^		요즘 인기 있는 '20대 여성 셔츠/블라우스'를 보여드리겠습니다.

그림 19. 의인화 조작 예시 [4]

(P38) 질문의 문구가 친근해서 좋았습니다. (P39) 친근함을 위해 존댓말이 아닌 반말로 질의가 진행되었던 점도 좋았던 거 같습니다. (P32) 구어체로 진행되어 설문이 딱딱하지 않고 재미있었습니다. 답변에 대해 오류가 있을 때 기다려 달라고 하는 말투 등도 좋았습니다.

개인화 또한 메신저를 사용하는 챗봇에서 제공해줄 수 있는 또 하나의 중요한 요소이다. 개인 정보나 구매 이력 등을 기반으로 추천서비스를 제공하는 수 많은 플랫폼들은 이미 대중화되어 있으며 개인화로 인해 제품이나 브랜드에 대해 호감도를 높이는 연구 결과도 많이 발표 되어 있다. [47, 48] 개인화를 적용하기 위해서 다양한 시도가 있었지만 사용자의 사용 이력 등의 축적된 데이터를 기반으로 개인에게 적당한 제품을 추천함으로써 그 효과를 낼 수 있겠지만 [25] 설문 조사의 경우에도 마찬가지로 여러 설문 조사를 진행한 이력이 존재해야 가능한 상황이 된다. 그 전에는 메신저에서 공개해주는 데이터를 기반으로 기본 정보를 가지고 개인화를 진행하는 것이 필요하다. 각각의

메신저에서 제공해주는 기본 정보의 차이는 있겠지만 해당 정보를 기반으로 이름을 불러 준다고나 성별에 따른 맞춤형 질문은 하는 등의 개인화를 만들어 내는 것이 필요하다. 하나의 계정으로 다수의 설문 조사를 진행하게 될 경우에는 축적된 데이터를 기반으로 관심이 있는 분야 등을 파악하여 설문 조사를 요청하여 그 응답률이나 응답 품질을 높일 수 있게 된다.

(P7) Chatbot 이 나를 기억 할 것 같은 느낌인데 그 점을 이용하여 내 정보를 앞으로도 계속 저장하고 있어 나의 모습을 한눈에 보여 줄 수 있는 화면을 보여주면 흥미 있어 하지 않을까 싶다. 예를 들어, 나의 취향을 분석해주거나. (P11) Chatbot의 프로필을 좋아하는 연예인으로 설정한다거나, 말투를 바꿔서 재미있게 하는 등 개인에게 맞춤 형으로 설문 하면 더욱 더 즐거워 질 것 같다. 예를 들어 사투리 버전이라던지, 요새 유행어를 많이 넣는다던지 등의 방법도 좋을 것 같다. (P15) 사용자가 관심이 있는 분야를 알아서 나중에 정보를 제공해주면서 업데이트 해준다면 다음 번에도 설문에 대한 참여도가 높아질 것 같다

## 개인정보 보호

일부 실험 참가자들은 메신저와 연동이 되어 있어 챗봇으로 설문조사를 진행하게 될 경우 개인 정보의 노출에 대한 문제점에 대해 언급하였다. 실제로 개인화와는 반대가 되는 개념이 될 수 있지만 개인 정보가 secure하게 보호되고 있다는 확신을 이용자에게 줘야 할 수도 있다. 필요할 경우 본인의 개인 정보가 연동되지 않도록 anonymous 로

설문 할 수 있는 옵션을 제공하는 것도 사용자로 하여금 본인이 선택할 수 있도록 해주는 것도 중요하다.

### 설문 진척상황 알림 및 오류 수정을 위한 디자인

실험 참가자들이 챗봇으로 설문 조사를 진행할 때 가장 불편한 점이 언제 끝날지 알 수가 없었다는 것이었다. 최근의 온라인 설문 조사들은 각 섹션 별로 얼마나 진행 되었는지를 쉽게 파악할 수 있었으나 대화형식의 챗봇에서는 이를 표현할 좋은 인터페이스는 없었다. 이로 인해 챗봇과의 대화를 통해 설문 조사하는 것은 재미이었지만 언제 끝날지에 대한 답답함은 존재한다고 하였다. 결국 대화 중간 혹은 새로운 인터페이스로 설문이 어느 정도 진행되었지 얼마나 남았는지를 주기적으로 알려줄 필요가 있다.

(P12) 설문이 언제 끝나는지 알았으면 좋겠습니다. 총 몇 가지 설문인지 포함하면 좋을 것 같네요. (P14) 질문의 양이 정해져 있다면 progress or %를 표시해 어느 정도 진행되었다는 것을 알 수 있으면 좋을 것 같습니다. (P34) 완료된 문항과 남아있는 문항에 대한 UI 추가가 필요해 보입니다. (P38) 설문은 무조건 길어질 경우 피로도가 늘어나므로 중간에 농담도 하고 설문이 어느 정도 진행되었는지에 대한 정보도 보여줘야 할 것 같습니다. (P37) 질문이 얼마나 남았는지 혹은 어느 정도 진행되었는지(%) 안보여서 답답했습니다.

그리고 실험으로 진행된 챗봇 설문조사의 경우 온라인 설문

조사와는 달리 일단 하나의 문항의 답변이 끝나게 되면 이를 수정할 수 없는 구조로 되어 있다. 이로 인해 사용자로부터 부정확한 답변을 받게 될 수 있기 때문에 이전 설문 조사 항목에 대한 수정을 할 수 있는 조치가 필요하다. 가령 “이전 답변 수정” 등의 명령을 받아 수정할 설문으로 돌아가고 해당 설문이 종료된 후 진행 중인 설문 문항으로 돌아오는 디자인을 통해 사용자가 느낄 수 있는 불편함을 해소할 수 있다.

### 반응 시간에 대한 합리적 디자인

마지막으로 각 방법 별 설문 조사 시간을 비교한 결과 챗봇의 설문시간이 온라인의 설문시간보다 더 오래 걸린 경우가 83.3%에 달했다. 1시간 이상 설문 조사가 진행 된 Case를 제외하고 챗봇 설문조사는 평균 21분 51초 (SD = 7분 1초)가 걸렸으며 온라인 설문 조사는 평균 16분 08초 (SD = 8분 18초)가 걸렸다. 실제로는 챗봇의 설문시간이 더 오래 걸렸으나 일부 실험 참가자들은 온라인 설문 조사가 더 오래 걸린 것 같다고 답변하기도 하였지만 실제 소요된 시간처럼 챗봇이 온라인 보다는 오래 걸린 것 같다는 답변이 65% 정도로 많이 차지했다. 결국 이미 정해진 질문, 보기들을 한번에 웹 페이지에 로드에서 진행하는 온라인 설문 조사에 비해 챗봇 설문 조사의 경우 대화형식인 관계로 매번 새로운 질문 및 보기를 보여줘야 하기 때문에 실험 참가자로 하여금 “기다린다”, “오래 걸린다” 라는 느낌을 주게 되었다. 실제로 실험 참가자가 응답한 후 대기 시간 없이 다음 질문을 보여줄 수 있도록 디자인을 할 수도 있지만 사용자 입장에서는 기계

적이라는 느낌을 받을 수 있어 질문 길이에 따라 1초 ~ 3초 정도의 gap을 추가하여 디자인을 한 것도 그 영향이 있다고 판단된다.

(P31) 챗봇의 경우 새로운 설문이 나오는데 약간의 갭이 있어서 좀 더 오래 걸릴 것 같다. (P4) Chatbot을 이용한 Survey가 오래 걸린다고 느껴졌습니다. 응답 후 기다려야 하며 Web Survey 때 속도를 못 따라왔습니다.

하지만, 대화 형식에 따른 기다림은 꼭 부정적인 영향을 주는 것은 아니고 설문을 좀 더 꼼꼼히 읽게 되는 긍정적인 영향도 있었으며 집중을 하게 되어 시간이 더 걸리는 것으로 생각하는 실험 참가자도 많았다. 결국 설문 조사의 길이를 사용자가 지루하다고 느껴지지 않을 만큼 짧게 가져가도록 설계하는 것이 필요할 것이다.

(P1) 체감상 Chatbot을 이용한 survey가 더 오래 걸리는 느낌은 있었다. Chatbot은 의견도 읽고, 대답을 좀 더 자세히 하게 되는 경향이 있는 것 같다. (P11) 챗봇의 설문시간이 더 오래 걸리는 것 같습니다. 하지만 설문을 좀 더 꼼꼼히 읽게 되는 장점은 있는 것 같습니다. (P14) 웹의 경우 대충/빨리 하고자 하였으나 챗봇 설문 조사의 경우 답변을 고민하여 전달하다 보니 오래 걸리는 것 같았다. (P21) Chatbot이 집중해서 읽고 답변하게 되므로 설문 시간이 더 오래 걸리는 것 같다. (P38) 챗봇의 설문 시간이 더 오래 걸리는 것 같지만 설문 문항이 작으면 더 짧다고 느낄 수도 있을 것 같다

## 제 5 장 결 론

### 제 1 절 연구 요약

본 연구에서는 챗봇이 설문조사를 위해 사용할 수 있는 feasibility가 있음을 정량적인 데이터를 기준으로 확인할 수 있었다. 특히 주관적인 의견을 묻는 주관식의 질문에 대해서는 챗봇 설문조사의 응답 품질이 훨씬 좋은 것으로 파악되었다. 질문 항목에 대한 무응답의 경우에 대해서는 주관적인 의견뿐만 아니라 객관적 사실에 대해서도 챗봇 설문조사가 온라인 설문조사에 비해 그 결과가 좋게 나타났으며 이는 대화형식이라는 챗봇 설문조사의 특징에 따른 효과라 볼 수 있다. 대부분의 연령대에서 익숙한 모바일 메신저라는 인터페이스를 사용함으로써 이를 통해 설문조사를 하는 것에 대한 거부감은 거의 없었으며 사용자들도 챗봇 설문 조사의 유용성에 대해서도 긍정적으로 나타났고 대화형 인터페이스가 주는 몰입감으로 인해 설문조사에 집중할 수 있게 되어 응답 품질도 더불어 좋아지게 되었다. 사람이 아닌 챗봇으로 1:1설문조사의 경험을 제공할 수 있게 됨으로써 비용을 줄이면서 Virtual Interviewer의 효과를 낼 수 있게 됨으로써 앞으로의 그 영역을 확장하게 되는 것은 쉽게 예측이 가능하다.

그리고 기존에 많이 사용되는 온라인 설문조사와의 비교를 통해

몇 가지 다른 점을 발견할 수 있었다. 대화형식을 사용하기 때문에 이에 대한 몰입감도 있어 대화 자체를 즐기면서 설문조사를 진행하는 경우가 많았다. 사람과의 대화와 유사하기 때문에 단순히 무시하거나 답변을 무성의하게 하지 않게 된다는 의견들도 많았다. 그리고 모바일 메신저를 이용함으로써 새로운 틀을 설치하거나 링크로 이동해서 설문을 하지 않아 접근성도 좋고 편하다는 장점이 있으며 기존에 자주 사용하고 있는 메신저의 경험의 연속성 및 특히 비대면의 형식을 가지고 있는 대화형식으로 인해 거부감도 낮게 나타나게 되었다. 중도 이탈자의 경우 온라인에서는 다시 설문조사를 처음부터 요청하거나 완전히 이탈하게 되지만 챗봇의 경우에는 질문이 끝난 시점부터 다시 자연스럽게 쉽게 이어나갈 수 있다. 이러한 특징들로 인해 설문조사에서 가장 중요하게 생각하는 응답률이 높게 나올 수 있다고 보여진다. 그리고 Open-ended Question에서의 답변의 길이가 더 길어지고 답변을 하지 않거나 거부하는 경우는 훨씬 줄어드는 것으로 나타나 응답품질도 증가되었다고 볼 수 있다. 이렇게 모바일 메신저를 자주 사용하는 사람들에게 설문이 필요할 경우에는 챗봇을 이용하는 것이 도움이 된다고 할 수 있다.

그리고 이 연구에서 진행된 사용자와의 인터뷰를 통해서 아래와 같이 챗봇으로 설문조사를 진행할 때 고려해야 할 사항들에 대해 파악할 수 있었다. 먼저, 말투가 사용자의 경험에 큰 영향을 미치는 것을 파악할 수 있었다. 사용자의 경우 메신저는 대부분 친한 사람들과의 대화 경험을 가지고 있기 때문에 부드럽고 친근한 말투로 대화를 이어나가게 되면 몰입감이 생겨 사용자로부터 더 많은 답변을 이끌어 낼 수 있다. 그리고 모바일 메신저를 사용하여 사용자의 정보를 획득할 수



있다면 개인 정보를 파악하여 각각의 개인에 맞는 말투나 주제에 대한 질문을 할 수 있게 된다. 자신이 관심이 있는 항목들에 대한 설문조사일 경우 그 답변의 품질은 훨씬 높아지게 된다. 위와 같이 챗봇 설문 조사에 의인화 및 개인화를 통해 디자인을 하는 것도 고려할 필요가 있다. 그리고, 대화형식의 장점을 살려 interaction을 통해 사용자에게 정보를 제공하거나 피드백을 제공한다면 그 경험을 바탕으로 사용자는 더욱 더 설문 조사에 더 시간을 할애할 수 있다. 일반적으로 온라인 설문조사에서는 Progress Bar를 보여주지만 챗봇에서는 그러한 방식이 자연스럽지는 않다. 그러므로 대화 중간에 몇 개나 남았는지 혹은 얼마나 진행되었는지를 알려주는 게 좋다. 그리고 질문이 길 경우에는 한 번에 모두 발화하지 않고 의미 있는 기준으로 잘라 여러 개의 대화문으로 진행하는 것이 좋다. 한번에 많은 텍스트가 전달 될 경우 사용자의 피로감을 높이고 이해도가 떨어지게 된다. 온라인 설문 조사 보다는 설문 시간이 길어질 수 있으므로 사용자가 지루하다고 느끼지 않도록 설문 길이를 조절하는 것이 사용자에게 부정적인 경험을 주지 않게 된다. 기존 답변을 수정할 수 있는 인터페이스를 고려하며 개인정보에 민감한 사용자를 위해 개인정보를 Secure하고 있다는 정보나 노출된 정보를 사용하지 않는 옵션을 제공해주는 것도 다양한 사용자의 기대를 만족하게 된다. 이러한 점들을 챗봇을 디자인할 때 염두 해 두면 응답품질은 좀 더 좋아지게 된다.

## 제 2 절 연구의 시사점 및 의의

급변하는 기술의 발전과 새로운 패러다임의 등장으로 우리가 관심을 많이 기울이지 못했던 부분에서 그 시너지효과가 나타날 수 있다. 특히 챗봇이 주로 동작하는 모바일 메신저는 현 시대를 살아가는 대부분의 사람들이 아주 많이 사용하는 도구 중의 하나로 이 사실 하나만으로도 굉장한 잠재력을 가지고 있게 된다. 이에 설문조사를 모바일 메신저에 적용하는 것은 사용의 편리함과 우수한 접근성의 장점을 모두 지닐 수 있게 되고 프로그램 된 대화형 인터페이스인 챗봇을 접목하여 사람을 대체하고 언제 어디서나 응답을 할 수 있는 이상적인 조합이 만들어지게 되는 것이다.

최근에는 챗봇 설문조사가 전통적인 온라인 설문조사와 비교하여 온라인 응답 품질이 좋아질 수 있는지에 대한 연구가 진행되었지만 대부분은 질문에 대한 관심은 없이 객관식이나 주관식은 문항을 실험 참가자에게 던져 그 응답에 대한 품질로 챗봇이 온라인 설문조사보다 응답 품질이 좋다는 결론을 내리고 있다. 하지만 이는 설문 조사가 어떠한 목적이나 어떠한 질문을 가지고 있는지에 대해서는 관심을 가지지 않은 것이 사실이다. 이에 본 연구에서는 질문의 내용에 대해 객관적 사실에 대한 질문인지, 주관적 의견을 묻는 질문인지에 대해 1차로 구분을 하고 설문에서 빠지지 않는 개인정보에 대해서는 따로 카테고리화 하여 설문을 진행하여 그 결과를 확인하는 단계를

거치게 되었다. 이를 통해서 설문 목적이나 질문의 유형에 따라 챗봇으로 설문조사를 하는 것이 더 좋을지 아니면 온라인 설문조사로도 유의차 없이 응답을 받을 수 있는지를 보고자 하였다.

본 연구의 결과에 따르면 주관적인 의견을 묻는 질문의 주관식에 대해서는 응답 길이와 더불어 No Response에 대해서도 챗봇의 결과가 온라인 설문조사의 결과보다 좋다는 것을 확인할 수 있었고 객관식에 대해서는 No Response에 대해서는 챗봇이 좋게 나타났지만 Differentiation에 대해서는 유의차가 없었다. 결국 주관적인 의견을 묻는 질문에 대해서는 챗봇으로 할 경우 객관식을 사용하는 것보다는 주관식 형태를 통해 사용자에게 직접 입력을 받도록 하는 것이 좀 더 좋은 응답을 받을 수 있는 것으로 판단할 수 있다. 객관적인 사실을 묻는 질문에 대해서는 주관식의 경우 응답 길이가 서로 유의차가 없는 것으로 보아 사실(Fact)에 대한 응답 자체는 사용자의 입장에서 대화형과는 관련없이 일정한 답변을 전달하게 되는 특징이 있어 객관적 사실에 대해서는 챗봇으로 설문 조사를 진행할 경우 주관식 형태의 질문 보다는 객관식 형태의 질문이 좀 더 나은 것으로 파악이 된다. 개인 정보에 대해서는 챗봇과 온라인 설문조사의 유의차가 없는 것으로 보면 다수의 실험 참가자가 사후 인터뷰에서 언급한대로 개인 정보에 대해서는 모든 사람이 민감하게 생각하고 그게 챗봇이든 어떤 다른 설문조사의 형태가 되더라도 가장 최소한의 응답을 하게 되는 특징을 가진다고 볼 수 있다. 결국 대화형 인터페이스라는 챗봇의 주요한 특징이 있더라도 많은 대화를 통한 인간적인 유대감을 맺기 전에는 개인 정보에 대한 응답품질은 크게 향상되지 않을 것으로 보여진다. 이는

챗봇이 설문 조사라는 단일의 활동을 위해서만 사용자와 대화를 하는 것이 아니라 지속적인 커뮤니케이션을 통해 유대관계를 이어나가고 이를 통해 상대방의 관심사에 대해 파악을 하고 정보를 제공하면서 사용자가 관심을 가질만한 설문조사를 진행할 수 있다면 그 설문의 응답 품질은 더욱 향상될 것으로 보인다.

위에서 논의한 것과 같이 본 연구를 통해 챗봇을 통한 설문조사에서 개인 정보에 대한 질문, 단순한 객관적 사실에 대한 질문, 본인의 주관적 의견에 대한 질문 등의 세분화된 질문 형태로 그 효과를 확인한 것이 본 연구에 의의가 있다.

그리고 모바일 메신저와 챗봇을 설문조사에 접목함으로써 발생할 수 있는 효과에 대해 학문적으로 탐색했다는데에도 의의가 있으며 이를 온라인 설문조사와의 비교 실험을 통해 정량적 데이터를 분석하고 사후 인터뷰를 통해 챗봇 설문 조사 방법이 가지는 여러가지 효과에 대해서도 확인하게 되어 앞으로 설문조사에서 챗봇의 시장이 확장될 수 있는데 이론적인 배경이 될 수 있다는 산업적인 의의가 있고 설문 조사 결과와 사후 인터뷰를 통해 얻어진 데이터를 기반으로 챗봇 설문 조사 디자인시 고려해야할 사항을 제안한 것에도 의의가 있다.

또한 본 연구는 설문조사라는 사회학적인 학문에 챗봇이라는 기술을 도입하여 연구를 진행했으며 이로 인해 설문조사의 결과를 향상시키는 등 주제 및 방법론에 있어서 융합적인 의미도 내포하고 있다고 할 수 있다.

### 제 3 절 연구의 한계 및 Future Work

본 연구의 그 첫 번째 한계점으로는 실험에서 발생하게 된 Selection Bias이다. 실험참가자가 20대부터 40대로 일부 세대의 인원으로 실험이 진행되어 전 연령의 데이터를 분석하지 못하여 그 결과를 일반화하기는 쉽지 않게 되었다. 추후에는 카카오톡 등을 이용하여 10대부터 노년층까지 모바일폰을 사용하는 모든 연령대를 포함한 실험으로 그 결과를 일반화 시킬 수 있도록 해야 하겠다. 특히 고령자의 경우 일반적인 웹 설문조사에는 익숙하지 않으나 카카오톡과 같은 SNS 챗봇에는 상당히 익숙해져 있기 때문에 대화를 통한 설문조사를 통해 기존 전화 설문조사를 대체할 수 있을만한 부분이고 이를 연구를 통해 증명할 수 있을 것이라 생각된다.

그리고 이 연구에서 사용된 챗봇은 Rule Based로 만들어져 사용자에게 대한 반응에 있어서 한계가 있게 되었다. 이미 정의된 형식의 구성과 내용들로 채워져 있어 개인화를 하지할 못하였고 사용자의 이탈행위에 대한 모든 대비는 이루어지지 못하였다. 만약 Generative Model의 챗봇을 제작하고 이를 설문조사에 사용함으로써 사용자들에게 꼬리에 꼬리를 무는 질문들도 할 수 있게 되고 좀 더 깊은 대화로 양질의 응답을 유도해 낼 수 있으리라 생각된다. 이 것은 사람이 아닌 챗봇이 스스로 생각하여 이용자와 대화하여 설문대화를 이어나가게 되어 기존의 전통적인 설문조사 방법을 모두 포괄할 수 있는 가장 이상적인

설문조사에 근접할 수 있다고 생각된다.

본 연구에서는 기본적으로 텍스트를 기반으로 하는 챗봇에 대한 연구를 진행했기 때문에 아마존의 Alexa나 네이버의 클로버 등의 음성인식 대화형 인터페이스에 대해서는 연구가 진행되지 못하였다. 설문조사에 대해서 음성 스피커를 이용하여 연구를 함으로써 다양한 방식의 대화형 에이전트를 통한 설문조사의 효과를 확인해 볼 수 있는 기회가 될 것이라 본다.

## 참 고 문 헌

- [1]. 이지영, & 김진교. (2006). 온라인 패널 조사와 오프라인 대인면접 조사 간 데이터 품질 비교: 신뢰도와 타당도를 중심으로.
- [2]. Kim, S., Lee, J., & Gweon, G. (2019, May). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–12).
- [3]. 백현지, 김상연, & 이상원. (2019). 챗봇 기반 인터페이스의 상호작용성과 사용 모드가 사용자 경험에 미치는 영향. 한국 HCI 학회 논문지, 14(1), 35–43.
- [4]. 정슬기. (2019). 패션 쇼핑 챗봇의 의인화와 개인화가 서비스 수용의도에 미치는 영향 (Doctoral dissertation, 서울대학교 대학원).
- [5]. 홍은지, 조광수, & 최준호. (2017). 스마트홈 대화형 인터페이스의 의인화 효과: 음성-채팅 인터랙션 유형에 따른 실험 연구. 한국 HCI 학회 논문지, 12(1), 15–23.
- [6]. 김현숙. (2013). 스마트폰 브랜드 선택에 있어서 소비자의 가치구조 규명에 관한 연구 수단: 목적 사슬이론을 중심으로. 광고연구, (97), 78–121.
- [7]. Olsen, F., Abelsen, B., & Olsen, J. A. (2012). Improving response rate and quality of survey data with a scratch lottery ticket incentive. BMC medical research methodology, 12(1), 52.
- [8]. Shawar, B. A., & Atwell, E. (2007, January). Chatbots: are they really useful?. In Ldv forum (Vol. 22, No. 1, pp. 29–49).
- [9]. Vicente, P., Reis, E., & Santos, M. (2009). Using

mobile phones for survey research: a comparison with fixed phones. *International Journal of Market Research*, 51(5), 1–16

- [10]. Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus PC web on survey response quality: A crossover experiment in a probability web panel. *Public Opinion Quarterly*, 81(S1), 280–306.
- [11]. Revilla, M., & Ochoa, C. (2015). What are the links in a Web survey among response time, quality, and auto-evaluation of the efforts done?. *Social Science Computer Review*, 33(1), 97–114.
- [12]. Saenz, J., Burgess, W., Gustitis, E., Mena, A., & Sasangohar, F. (2017). The usability analysis of chatbot technologies for internal personnel communications. In *IIE Annual Conference. Proceedings* (pp. 1357–1362). Institute of Industrial and Systems Engineers (IISE).
- [13]. Banda, J. P. (2003, December). Nonsampling errors in surveys. In *Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys*. United Nations Secretariat, New York (pp. 3–5).
- [14]. Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- [15]. Galvao, A. M., Barros, F. A., Neves, A. M., & Ramalho, G. L. (2004, July). Persona-AIML: an architecture for developing chatterbots with personality. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004.* (pp. 1266–1267). IEEE.
- [16]. Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology* (Vol. 561). John Wiley & Sons.
- [17]. Evans, J. R., & Mathur, A. (2005). The value of online



surveys. Internet research.

- [18]. Ferber, R., Sheatsley, P., Turner, A. G., & Waksberg, J. (1980). What is a Survey?. EMBRAPA–DMQ.
- [19]. Guan, J., & Alkinkemer, K. (2002). Instant messaging: Chatting with your customers online and beyond. AMCIS 2002 Proceedings, 67.
- [20]. Bosnjak, M., Poggio, T., Becker, K. R., Funke, F., Wachenfeld, A., & Fischer, B. (2013). Online survey participation via mobile devices. Retrieved 29th of July.
- [21]. Wenz, A. (2017). Completing web surveys on mobile devices: does screen size affect data quality? (No. 2017–05). ISER Working Paper Series.
- [22]. de Bruijne, M. (2015). Designing web surveys for the multi–device internet. Tilberg: Center for Economic Research. Available at: [https://pure.uvt.nl/ws/files/8728830/Thesis\\_MarikadeBruijne.pdf](https://pure.uvt.nl/ws/files/8728830/Thesis_MarikadeBruijne.pdf) [Accessed 12/12/18].
- [23]. Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public opinion quarterly*, 68(3), 368–393
- [24]. Bannan, K. J. (2003). Companies save time, money with online surveys. *B to B*, 88(6), 1–2.
- [25]. Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, 106359.
- [26]. Mavletova, A., & Couper, M. P. (2014). Mobile web survey design: scrolling versus paging, SMS versus e–mail invitations. *Journal of Survey Statistics and Methodology*, 2(4), 498–518.

- [27]. Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior research methods*, 45(2), 527–535.
- [28]. Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of computer-mediated communication*, 6(2), JCMC621.
- [29]. Sauermann, H., & Roach, M. (2013). Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Research Policy*, 42(1), 273–286.
- [30]. Trespalacios, J. H., & Perkins, R. A. (2016). Effects of personalization and invitation email length on web-based survey response rates. *TechTrends*, 60(4), 330–335.
- [31]. Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, 60(6), 821–836.
- [32]. Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public opinion quarterly*, 68(1), 94–101.
- [33]. Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, 73(2), 349–360.
- [34]. Yu, S., Alper, H. E., Nguyen, A. M., Brackbill, R. M., Turner, L., Walker, D. J., ... & Zweig, K. C. (2017). The effectiveness of a monetary incentive offer on survey response rates and response completeness in a longitudinal study. *BMC medical research methodology*, 17(1), 77.
- [35]. Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International*

journal of market research, 50(1), 79–104.

- [36]. Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69(3), 370–392.
- [37]. De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of official statistics*, 21(5), 233–255.
- [38]. De Leeuw, E. D., & Hox, J. J. (2011). Internet surveys as part of a mixed-mode design. *Social and behavioral research and the Internet*.
- [39]. Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3), 439–457.
- [40]. Wagner, J. R. (2008). *Adaptive Survey Design to Reduce Nonresponse Bias* (Doctoral dissertation).
- [41]. Tourangeau, R., Michael Brick, J., Lohr, S., & Li, J. (2017). Adaptive and responsive survey designs: A review and assessment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 203–223.
- [42]. Cahn, J. (2017). *CHATBOT: Architecture, design, & development*. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.
- [43]. Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9), 641–661.
- [44]. Lund, A. (2001). Measuring usability with the USE Questionnaire. *Usability Interface: The usability SIG newsletter of the Society for Technical Communications*, 8.
- [45]. Dillman, D. A. (2011). *Mail and Internet surveys: The*

tailored design method—2007 Update with new Internet, visual, and mixed-mode guide. John Wiley & Sons.

- [46]. Archer, T. M. (2008). Response rates to expect from web-based surveys and what to do about it. *Journal of extension*, 46(3), 1–5.
- [47]. Kang, M., Shin, D. H., & Gong, T. (2016). The role of personalization, engagement, and trust in online communities. *Information Technology & People*.
- [48]. Ho, S. Y., & Bodoff, D. (2014). The effects of Web personalization on user attitude and behavior. *MIS quarterly*, 38(2), 497–A10.
- [49]. Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., ... & Matsuo, Y. (2014, August). Towards an open-domain conversational system fully based on natural language processing. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 928–939).
- [50]. Schonlau, M., Ronald Jr, D., & Elliott, M. N. (2002). Conducting research surveys via e-mail and the web. Rand Corporation.
- [51]. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611.
- [52]. Cobanoglu, C., Moreo, P. J., & Warde, B. (2001). A comparison of mail, fax and web-based survey methods. *International journal of market research*, 43(4), 1–15.
- [53]. Kwak, N., & Radler, B. (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of official statistics*, 18(2), 257.
- [54]. Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public opinion quarterly*, 68(1), 94–101.

- [55]. Scholl, N., Mulders, S., & Drent, R. (2002). On-line qualitative market research: interviewing the world at a fingertip. *Qualitative Market Research: An International Journal*.
- [56]. Curasi, C. F. (2001). A Critical Exploration of Face-to Face Interviewing vs. Computer-Mediated Interviewing. *International Journal of Market Research*, 43(4), 1–13.
- [57]. Haslam, N. (2006). Dehumanization: An integrative review. *Personality and social psychology review*, 10(3), 252–264.
- [58]. Shapiro, M. J. (1970). Discovering interviewer bias in open-ended survey responses. *The Public Opinion Quarterly*, 34(3), 412–415.
- [59]. De Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). Mixed-mode surveys: when and why. *International handbook of survey methodology*, 299–316.
- [60]. Rossi Peter, H., Wright James, D., & Anderson Andy, B. (1983). *Handbook of survey research* Orlando.
- [61]. Ilieva, J., Baron, S., & Healey, N. M. (2002). Online surveys in marketing research. *International Journal of Market Research*, 44(3), 1–14.
- [62]. Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(1), 51–68.
- [63]. Goetz, E. G., Tyler, T. R., & Cook, F. L. (1984). Promised incentives in media research: a look at data quality, sample representativeness, and response rate. *Journal of Marketing Research*, 21(2), 148–154.
- [64]. Dillman, D. A. (1991). The design and administration of mail surveys. *Annual review of sociology*, 17(1), 225–249.

- [65]. Templeton, L., Deehan, A., Taylor, C., Drummond, C., & Strang, J. (1997). Surveying general practitioners: does a low response rate matter?. *Br J Gen Pract*, 47(415), 91–94.
- [66]. Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). John Wiley & Sons.
- [67]. Groves, R. M. (1987). *Research on Survey Data Quality Public Opinion. Quality by*, 51, 3.
- [68]. Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing letters*, 15(1), 21–36.
- [69]. McCarty, J. A., & Shrum, L. J. (2000). The measurement of personal values in survey research: A test of alternative rating procedures. *Public Opinion Quarterly*, 64(3), 271–298.
- [70]. Shechtman, N., & Horowitz, L. M. (2003, April). Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281–288).
- [71]. Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice*, 12(2).
- [72]. Miles, J., & Gilbert, P. (Eds.). (2005). *A handbook of research methods for clinical and health psychology*. Oxford University Press on Demand.
- [73]. Matjašić, M., Vehovar, V., & Manfreda, K. (2018). Web survey paradata on response time outliers: A systematic literature review. *Metodološki zvezki*, 15(1), 23–41.
- [74]. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*,

3(2), 77–101.

- [75]. Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. sage.
- [76]. Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test* (pp. 23–65). Springer, Dordrecht.
- [77]. Ackerman, E. (2014). A better test than turing [News]. *IEEE Spectrum*, 51(10), 20–21.
- [78]. Gigliotti, L., & Dietsch, A. (2014). Does age matter? The influence of age on response rates in a mixed-mode survey. *Human Dimensions of Wildlife*, 19(3), 280–287.
- [79]. Wambsganss, T., Winkler, R., Söllner, M., & Leimeister, J. M. (2020, April). A Conversational Agent to Improve Response Quality in Course Evaluations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–9).
- [80]. Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2019). Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys. *arXiv preprint arXiv:1905.10700*.
- [81]. Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public opinion quarterly*, 72(5), 836–846.
- [82]. Jones, T. L., Baxter, M. A. J., & Khanduja, V. (2013). A quick guide to survey research. *The Annals of The Royal College of Surgeons of England*, 95(1), 5–7.

## Abstract

# Comparative analysis of response quality according to the questionnaire of chatbot and online survey:

Focusing on objective facts, subjective opinions, and personal information questions

Beomgeun Bae

Program in Digital Contents and Information Studies

Department of Transdisciplinary Studies

The Graduate School

Seoul National University

Surveys are one of the important ways to understand the nature of a population. In recent years, online surveys are most frequently used, but they still suffer from low response rates or poor response quality. In this paper, as a new method of survey, we study whether chatbots are useful according to the types of questions such as subjective opinion, objective facts, and personal information, and discuss the



differences between surveys using chatbots and surveys using online.

To achieve this research goal, we created a chatbot prototype for the survey using the iterative design process to improve the design through feedback received through several experiments. The chatbot was created using the fluently.ai platform and the Chatfuel platform and was designed to operate based on Facebook messenger.

Through several user studies, it was found that factors related to Human Nature, such as soft speech and emotional sympathy, contributed positively to the results of the survey. These characteristics are defined as the main elements of the chatbot, and based on quantitative results for “No Response” , “Response Length” , and “Differentiation” of responses by question through comparative experiments with online surveys. It was confirmed that the chatbot survey differed in the quality of responses in objective facts and subjective opinions than online survey. A qualitative analysis of chatbot survey was also conducted through post-interviews. As a result, the engagement and usefulness of the survey was confirmed through the chatbot.

This study has provided a theoretical foundation to show that chatbots can be used as an important methodology for surveys, and the results can be effective. For surveys that require subjective opinion, chatbots are more effective than online surveys. It is also meaningful to present some design considerations for reference to

apply the chatbots identified through several experiments to the survey.

Keywords: Chatbot, Survey, Response Quality, Subjective Opinion, Objective Fact

Student Number: 2017-26601